


```

Db 1  MPMLPWVAEVRRLSGQCSKKEPRTKQVRLWVSPSGLRCEPDLEKSPWDPLICSSIFEC 60
Qy 97  KQORVHKLJHNSHDPYSFACILKEDAVHROSQCYVPKADDTQKVPEIISIRQAGKIARQ 156
Db 61  KQORVHKLJHNSHDPYSFACILKEDAAHROSQCYVPKADDTQKVPEIISIRQAGKIARQ 120
Qy 157  EELHCPSEFDDTFSKFEVLFQGRVTVAHKAPPALIDECIEKFNHVSGRSSESPRNP 216
Db 121  EELHCPSEFDDTFAKFEVLFQGRVTVAHKAPPALIDECIEKFNHVSGRRTD----- 174
Qy 217  PHAAPTGS--QEPVRRPMRKSQFQGLRSLAFRAKELQDGLRSGFFSFEESDIENHLIS 275
Db 175  -WEAPTGPSAPGPRPMRKSQFQGLRSLAFRAKELQDGLRSG--TFSSP-DNDIENHLIG 231
Qy 276  GNIVQPTDIENRMLFTIGOSEVYVLIISPDTKKTALEKNFEKISFCOGIRHVDHFGPI 335
Db 232  GNIVQPTDMEENRMLFTIGSEVYVLIISPDTKKTALEKNFEKISFCOGIRHVDHFGPI 291
Qy 336  CRESSGG--GGHFVFCVQCTNEALVDEIMMTLKQAFVAAVQQTAKAPALCEGCPILQ 393
Db 292  CRECSGGSGGFHFVFCVQCTNEALVDEIMMTLKQAFVAAVQQTAKAPALCEGCPILQ 351
Qy 394  SLHKLCERTEGNNSKTKLELOKHLTTLNQEQATIFEVQKLRPRNEORENELIISFLR 453
Db 352  GLHKLCERTEGNNSKTKLELOKHLTTLNQEQATIFEVQKLRPRNEORENELIISFLR 411
Qy 454  CLYEKOKEHIGHGEMKQTSMAAENIGSELPPSATFRFLDMLKNKAKRSLTESLESILS 513
Db 412  CLYEKOKESHGTGPKOTQVAEENIGSDLPSPASRFLDSLKNKAKRSLTESLESILS 471
Qy 514  RGNKARGLOHESISVDLSSLSSTLSNTSKEPSVCEKEALPSESFKLLGSSDLSDDS 573
Db 472  RGNKARGLOHESISVDLSSLSSTLSNTSKELSMGDKAEPVSETSFKLLGSSDLSDDS 531
Qy 574  ESHLPEEPAPLSPQAFRRRANTLSHPFIECOEPPQARGSPGVQKLMRYHSVSTETPP 633
Db 532  EGHIAEESALLSPQAFRRRANTLSHPFIECPAPPEPAQSPGVQKLMRYHSVSTETPP 591
Qy 634  HERKDESKANHLGDSGPGVTRHRSWROQIFLRVATPQKACDSSRYEDYSELGELPP 693
Db 592  HERKDESKANHLGDSGPGVTRHRSWROQIFLRVATPQKACDSSRYEDYSELGELPP 651
Qy 694  RSPLEPCVCDGPGFPPEPEKRTSRELWQKAILQOITLLRMKEKNOQLQASENDLIN 753
Db 652  RSPLEPCVCDGPGFYKRRGRHASFESCGRRPSCSRSLYRMEKNOQLQASENDLIN 711
Qy 754  KRLKLDYEITPCLKEVTVWEKMLSTPGRSKIKFDMKMKHSAVGGVPRHRHGEIWKPL 813
Db 712  KRLKLDYEITPCLKEVTVWEKMLSTPGRSKIKFDMKMKHSAVGGVPRHRHGEIWKPL 771
Qy 814  AEQFHLKHOFPSKQOPKQVPYKELLKQTSQOHAILEDLGRFTPTHPYFSAOLGAGQSL 873
Db 772  AEQFHLKHOFPSKQOPKQVPYKELLKQTSQOHAILEDLGRFTPTHPYFSAOLGAGQSL 831
Qy 874  YNLKAYSLLDQEVGYCOGLSFVAGILLHMESEEAFAKMLKFLMDFMDGLRKQYRPMIIL 933
Db 832  YNLKAYSLLDQEVGYCOGLSFVAGILLHMESEEAFAKMLKFLMDFMDGLRKQYRPMIIL 891
Qy 934  QIOMYQLSRLLDYHRLDNLHLEEHIGPSLYAAPWFLTMFASQPLGFAVRFDMIFLQ 993
Db 892  QIOMYQLSRLLDYHRLDNLHLEEHETGPTTAAAPWFLTMFASQPLGFAVRFDMIFLQ 951
Qy 994  GTEVIFKVALSLGSHKPLILOHENLETIVDTFKSTPLNGLVQMEKTNQVFMEDIAKQ 1053
Db 952  GTEVIFKVALSLGSHKPLILOHENLETIVDTFKSTPLNGLVQMEKTNQVFMEDIAKQ 1011
Qy 1054  LQAEVEYHVLQEEILDSPLSDNORMDKLEKTNLSLRKQNLDDLEQLQVANGRIQSLEA 1113
Db 1012  LQAEVEYHVLQEEILDSPLSDNORMDKLEKTNLSLRKQNLDDLEQLQVANGRIQSLEA 1071
Qy 1114  TIEKLSSSESKLKQAMLTLELER 1136
Db 1072  TVEKLLTSSESKLKQALTTLEVER 1094

```

```

RESULT 2
Q9UPP4 ID Q9UPP4 PRELIMINARY; PRT: 763 AA.
AC Q9UPP4;
DC 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE KIAA1108 PROTEIN (FRAGMENT).
GN KIAA1108.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno A., Nagase T., Ishikawa K., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RL DNA Res. 6:197-205(1999).
DR EMBL: AB029031; BAA83060.1;
DR InterPro: IPR001220; Lectin_legB.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00164; TBC; 1.
DR PROSITE: PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 763 AA; 87585 MW; 32F7C3E2E3A49A83 CRC64;

```

```

Query Match 64.6%; Score 3913; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 2.4e-250;
Matches 763; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 406 NSSKTKLEQLKHLTTLNQEQATIFEVQKLRPRNEORENELIISFLRCLYEKQKEH 465
Db 1 NSSKTKLEQLKHLTTLNQEQATIFEVQKLRPRNEORENELIISFLRCLYEKQKEH 60
Qy 466 IGEKMTQSMAAENIGSELPPSATFRFLDMLKNKAKRSLTESILSRGNKARGLQEH 525
Db 61 IGEKMTQSMAAENIGSELPPSATFRFLDMLKNKAKRSLTESILSRGNKARGLQEH 120
Qy 526 ISVDLSSLSSTLSNTSKEPSVCEKEALPSESFKLLGSSDLSSESHPLEPAPL 585
Db 121 ISVDLSSLSSTLSNTSKEPSVCEKEALPSESFKLLGSSDLSSESHPLEPAPL 180
Qy 586 PQQAFRRRANTLSHPFIECOEPPQARGSPGVQKLMRYHSVSTETPHERKDFESKAN 645
Db 181 PQQAFRRRANTLSHPFIECOEPPQARGSPGVQKLMRYHSVSTETPHERKDFESKAN 240
Qy 646 LGDSGTPVKTTRHRSWROQIFLRVATPQKACDSSRYEDYSELGELPPRSPLEPCD 705
Db 241 LGDSGTPVKTTRHRSWROQIFLRVATPQKACDSSRYEDYSELGELPPRSPLEPCD 300
Qy 706 FGPPPEEKRTSRELWQKAILQOITLLRMKEKNOQLQASENDLKNRKLDEEITP 765
Db 301 FGPPPEEKRTSRELWQKAILQOITLLRMKEKNOQLQASENDLKNRKLDEEITP 360
Qy 766 CLKEVTVWEKMLSTPGRSKIKFDMKMKHSAVGGVPRHRHGEIWKFLAEQFLKHQPP 825
Db 361 CLKEVTVWEKMLSTPGRSKIKFDMKMKHSAVGGVPRHRHGEIWKFLAEQFLKHQPP 420
Qy 826 KOQKDVYPYKELLKQTSQOHAILEDLGRFTPTHPYFSAOLGAGQSLYNILKAYSLL 885
Db 421 KOQKDVYPYKELLKQTSQOHAILEDLGRFTPTHPYFSAOLGAGQSLYNILKAYSLL 480
Qy 886 EVGYCOGLSFVAGILLHMESEEAFAKMLKFLMDFMDGLRKQYRPMIILQIQYLS 945

```

```

Db 481 EVGYCOGLSFVAGILLHLMSEEEAFKMLKFLMDFMGLRKQYRPMIILQIOMYQLSRLH 540
QY 946 DYHDLNHLHEEHEIGFSLYAAPFLMTFASQFFLGFVARVDFMIFLOGTEVIFKVALS 1005
Db 541 DYHDLNHLHEEHEIGFSLYAAPFLMTFASQFFLGFVARVDFMIFLOGTEVIFKVALS 600
QY 1006 LGSHPILLOHENLETIVDFTKSLPNLGLVQMEKTTINQFEMDIKQIQAQYEVYHVQ 1065
Db 601 LGSHPILLOHENLETIVDFTKSLPNLGLVQMEKTTINQFEMDIKQIQAQYEVYHVQ 660
QY 1066 EELIDSSPLSDNQMDKLEKTNSSLRKQNDLLEQLQVANGRIQSLSEATIEKLLSSSKL 1125
Db 661 EELIDSSPLSDNQMDKLEKTNSSLRKQNDLLEQLQVANGRIQSLSEATIEKLLSSSKL 720
QY 1126 KQAMLTLELERSALLQVVEELRRRSAPSDREPECTQPEPTGD 1168
Db 721 KQAMLTLELERSALLQVVEELRRRSAPSDREPECTQPEPTGD 763

RESULT 3
Q96K82 PRELIMINARY; PRT; 674 AA.
ID Q96K82
AC Q96K82;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CNA FLJ1449 FIS, CLONE HEMB1001684.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBryo;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Nimomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027355; BAB55057.1;
SQ SEQUENCE 674 AA; 77151 MW; 095584FB3C800A58 CRC64;

Query Match 57.2%; Score 3463; DB 4; Length 674;
Best Local Similarity 100.0%; Pred. No. 1.2e-220;
Matches 674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 MLKNAKRSLTESLISRGNKARGLOEHSISVDLSLSSTLSNTSKPSVCEKEALP 554
Db 1 MLKNAKRSLTESLISRGNKARGLOEHSISVDLSLSSTLSNTSKPSVCEKEALP 60
QY 555 ISESSFKLLGSSDELSDSHSLPEEPAPLSPQAFRRRANTLSHFPIEQEPPQARGS 614
Db 61 ISESSFKLLGSSDELSDSHSLPEEPAPLSPQAFRRRANTLSHFPIEQEPPQARGS 120
QY 615 PGVSQRKLMRYHSYSTPHERKDFESKANHLGDSGGTPVKTRRHSWRQIFLRVATPQK 674
Db 121 PGVSQRKLMRYHSYSTPHERKDFESKANHLGDSGGTPVKTRRHSWRQIFLRVATPQK 180
QY 675 ACSSSSRYEDYSELGELPPRSPLEPVCEGDPFGPPPEKRTSRELWQKAILQOILL 734
Db 181 ACSSSSRYEDYSELGELPPRSPLEPVCEGDPFGPPPEKRTSRELWQKAILQOILL 240
QY 735 LRMEKENQKLOASNDLNLKRLDYEEITPCLKEVTVMKMLSTPGRSKIKFDMKMH 794
Db 241 LRMEKENQKLOASNDLNLKRLDYEEITPCLKEVTVMKMLSTPGRSKIKFDMKMH 300
QY 795 SAVGQGVPRHHRGIEWKFLAQFHLKHQFPKQKDPYKELLKOLTISOQHAILIDLGR 854

```

```

Db 301 SAVGQGVPRHHRGIEWKFLAQFHLKHQFPKQKDPYKELLKOLTISOQHAILIDLGR 360
QY 855 TFPTHYPFSAQLGAGQSLYNILKAYSLLDQEVGYCGGLSFVAGILLHLMSEEEAFKMLK 914
Db 361 TFPTHYPFSAQLGAGQSLYNILKAYSLLDQEVGYCGGLSFVAGILLHLMSEEEAFKMLK 420
QY 915 FLFMDGLRKQYRPMIILQIOMYQLSRLHLDYHRLYNHLEHEIGFSLYAAPFLMTF 974
Db 421 FLFMDGLRKQYRPMIILQIOMYQLSRLHLDYHRLYNHLEHEIGFSLYAAPFLMTF 480
QY 975 ASOFPLGFVARVDFMIFLOGTEVIFKVALSLLGSHKPLILQHENLETIVDFIKSTLPNLG 1034
Db 481 ASOFPLGFVARVDFMIFLOGTEVIFKVALSLLGSHKPLILQHENLETIVDFIKSTLPNLG 540
QY 1035 LVOMEKTTINQFEMDIKQIQAQYEVYHVQLEELIDSSPLSDNQMDKLEKTNSSLRKQ 1094
Db 541 LVOMEKTTINQFEMDIKQIQAQYEVYHVQLEELIDSSPLSDNQMDKLEKTNSSLRKQ 600
QY 1095 LDLEQLQVANGRIQSLSEATIEKLLSSSKLKQAMLTLELERSALLQVVEELRRRSAP 1154
Db 601 LDLEQLQVANGRIQSLSEATIEKLLSSSKLKQAMLTLELERSALLQVVEELRRRSAP 660
QY 1155 DREPECTQPEPTGD 1168
Db 661 DREPECTQPEPTGD 674

RESULT 4
O60343 PRELIMINARY; PRT; 1299 AA.
ID O60343
AC O60343;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KIAA0603 PROTEIN.
GN KIAA0603.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Onara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR EMBL: AB011175; BAA25529.1;
DR InterPro: IPR000050; PID_domain.
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam: PF00640; PID; 2.
DR Pfam: PF00566; TBC; 1.
DR SMART: SM00462; PTB; 2.
DR SMART: SM00164; TBC; 1.
DR PROSITE: PS01179; PID; 1.
SQ SEQUENCE 1299 AA; 146606 MW; 6EECC59BC62EEA88 CRC64;

Query Match 46.7%; Score 2828.5; DB 4; Length 1299;
Best Local Similarity 48.3%; Pred. No. 2.7e-178;
Matches 621; Conservative 184; Mismatches 291; Indels 191; Gaps 28;

QY 21 FGQLVGLSPVHSULTMPMPWVAEVRRLSRQSTR---KEPVTQVRLCVSPSGLRCEP 77
Db 36 FRLWYVGGSCLDHRTLPMLPWLMAEIRRRSQPEAGCGCAPAEVILVLSAPFLRCVP 95
QY 78 EPG-----RSQMDPLIYSIFECKPQRVHKLIHNSHDSYFACLTK--EDAVHRQ 126
Db 96 APCAGASGGTSPSATQNPVAF--IFEKHAHISRFTHNSHDLTYFAYLIKAQPDPEQ 153

```



```
Db 424 VOMEKTSIQVETDISKQQAQVEYEHVQLQELIDSSPLSDNQMDKLEKTNSSLRKQNL 483
Qy 1096 DLLSQLVANGRIQSLQATIEKLLSSSKLKQAMLTLELERSALLQVTEELRRRSAPSD 1155
Db 484 DLLSQLVANGRIQSLQATIEKLLSSSKLKQATLALERSALLQVTEELRRRSAPSD 543
Qy 1156 REPECTQPEPTGD 1168
Db 544 QESDPTLPKPSGD 556

RESULT 6
Q923F8 PRELIMINARY; PRT; 494 AA.
ID Q923F8
AC Q923F8
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE UNKNOWN (PROTEIN FOR IMAGE: 3500261) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Chordata; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004675; AA004675.1;
FT NON_TER 1
SQ SEQUENCE 494 AA; 57041 MW; 4C1E7856F84A5F4F CRC64;

Query Match 39.7%; Score 2404; DB 11; Length 494;
Best Local Similarity 94.1%; Pred. No. 8e-151;
Matches 465; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

Qy 675 ACDSRSSVEYSELGELPPSPLEPVCEGDFGPPPEKRTSRELRLQKATLQILL 734
Db 1 ACDSPSRYEDSELGELPPSPLEPVCEGDFGPPVQBEKRTSRELRLQKATLQILL 60
Qy 735 LRMEKENKQLQASNDLNRKLDYDEITPCLKEVTVVMEKMLSTPGRSKIKFDMKMH 794
Db 61 LRMEKENKQLQASNDLNRKLDYDEITPCLKEVTVVMEKMLSTPGRSKIKFDMKMH 120
Qy 795 SAVGQGVPRHRRGEIWKFLAEQFHLKHQFPKQKQDPVYKELKQTSQOHALIDLGR 854
Db 121 SAVGQGVPRHRRGEIWKFLAEQFHLKHQFPKQKQDPVYKELKQTSQOHALIDLGR 180
Qy 855 TFPHPHYSAQAGQSLNLYLKAYSLDQEVGYCOGLSFVAGILLHMSSEAFKMLK 914
Db 181 TFPHPHYSAQAGQSLNLYLKAYSLDQEVGYCOGLSFVAGILLHMSSEAFKMLK 240
Qy 915 FLFMDGLRKQYRPMIILQIQVLSRLLDYHRLYHLYNLEHEIGPSLYAAPFLTMF 974
Db 241 FLFMDGLRKQYRPMIILQIQVLSRLLDYHRLYHLYNLEHEIGPSLYAAPFLTMF 300
Qy 975 ASQPLGFVAVFDMIFLQTEVIFKVALSLGSHKPLLOHQENLETIVDFIKSTLNLG 1034
Db 301 ASQPLGFVAVFDMIFLQTEVIFKVALSLGSHKPLLOHQENLETIVDFIKSTLNLG 360
Qy 1035 LVOMEKTIQVFNENDIAKQQAQVEYEHVQLQELIDSSPLSDNQMDKLEKTNSSLRKQ 1094
Db 361 LVOMEKTIQVFNENDIAKQQAQVEYEHVQLQELIDSSPLSDNQMDKLEKTNSSLRKQ 420
Qy 1095 LDLLEQLQVANGRIQSLQATIEKLLSSSKLKQAMLTLELERSALLQVTEELRRRSAPSD 1154
Db 421 LDLLEQLQVANGRIQSLQATIEKLLSSSKLKQAMLTLELERSALLQVTEELRRRSAPSD 480
Qy 1155 DRPECTQPEPTGD 1168
Db 481 TPEPDCQTEPTGD 494

Query Match 24.2%; Score 1467.5; DB 5; Length 1379;
Best Local Similarity 32.1%; Pred. No. 2.9e-88;
Matches 372; Conservative 184; Mismatches 334; Indels 269; Gaps 32;

Qy 141 PEIISIRQAGK-----IARQELHCPSEFDDTFKSKFVLCGRVT 182
Db 67 PNPVSSKMKASKSYTHGLSSSGTGVNTPTSTSAQNSLLADISPNTHTFFVYVVGKIR 126
```

```
RESULT 7
Q9VNG9 PRELIMINARY; PRT; 1379 AA.
ID Q9VNG9
AC Q9VNG9
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PLX PROTEIN.
GN PLX OR CG1093.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RA SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "the genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003602; AAF51967.1;
DR FlyBase; FBgn004879; plx.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00556; TBC; 1.
DR SMART; SM00462; TBC; 1.
DR SMART; SM00164; TBC; 1.
SQ SEQUENCE 1379 AA; 154333 MW; 31DD6BE68CF4790E CRC64;
```

```
Query Match 24.2%; Score 1467.5; DB 5; Length 1379;
Best Local Similarity 32.1%; Pred. No. 2.9e-88;
Matches 372; Conservative 184; Mismatches 334; Indels 269; Gaps 32;

Qy 141 PEIISIRQAGK-----IARQELHCPSEFDDTFKSKFVLCGRVT 182
Db 67 PNPVSSKMKASKSYTHGLSSSGTGVNTPTSTSAQNSLLADISPNTHTFFVYVVGKIR 126
```

```
QY 183 VAHKAPALIDECIEKFN----- 201
Db 127 VSKRVPTFIDALPKFKAYDAQRLLQNRKMSLSSEGGVIEAKPSSLSKSHDLKEE 186
QY 202 -----HVSGRSGSEPPNPHPHAAPTGSOEPPVRRPMRKSF 236
Db 187 DEEEQOHKHGHDSDSOAKPLVOLQLTGAEBGAAPRPLEDNKE---NKSPEKPLLRQ 243
QY 237 SQ-----PGLSLAFRKELO-----DGLRSSGFFSFEESDIE-NH 272
Db 244 SOELGHKEHSDGSOPLAANSOLEAPNVIVNQPTPPRDOGVGTGTASAGPSQLHPNY 303
QY 273 LSGHNIVOPTD-----IEENRTMLFTIGOSEVYLISPDTKKIALEKNEKEISF 321
Db 304 AMD--NIPQRDSASQGIPIVVEQNRMTWFLVGRCDLRLSPDRKQVLLYKDFKDVAS 361
QY 322 CSQGIHVDFHFGICRESSGGGFHVVCYVFOCTNEALVDEIMMTLKQAFVAAVQQTAK 381
Db 362 CVHGKSLDHFGIICRELNDG---YIGVFKQSEHVCDIVAATAQAFTCA-EQKK 417
QY 382 APAQL--CEGCPLOSLLHKCEIEGNSSKTKLEQKHITTTNOEQATIFEVOKLRPR 439
Db 418 QDTQIFSCHECHPMLWYHKLTQVGLSEKKTQALIRRIETLSDDQEITVWAKFCGSEKT 477
QY 440 NE--ORENELIISFLRCLIEEKKHIGEMKOTSOMAAENIGSELPPSATRFLDMLK 497
Db 478 NSPVAEQNFLMMLRAHCESSQORHVH-----DTAENRSEFLNQYLGST-----IF 525
QY 498 NKAKRSLTESLSISRGKNGKRGLOEHSISVDLSLSTLNTSKPSCVCEKALPISE 557
Db 526 MKAKRSLTNSFNLLKR-----KPS---KDDIAVPS 553
QY 558 SEFKLL--CSSEDLSDSSESHLPEEPAPLSPOAARRRANTLSHPICEQPPQARGSP 615
Db 554 HNLDRIREGSAEPPLGTQ-----SPPEGRFSNTVTGASP-----SSKP 591
QY 616 GVSQRK--LMRVHSVSTETPHERKOPESKANHLGDSGGTPVKTRRHSWRQOIFLRVATPQ 673
Db 592 TAEQLKSPMDIFIKVGNPK-----AETHQGSWQAILNSVVTPS 633
QY 674 KACDSSRYEDISELGEPPRPLEPVCEDGPPGPPPEKKTSLRELRLWOKAILQQL 733
Db 634 KGLDSEVPTFLS-----PMRPAKRGKRDAAELRELWRTAINQTIM 675
QY 734 LRMKEKOKLOASENDLLKRLKLDYEITPCLKEVTVWEKML---STP-GRSKIKFD 789
Db 676 LRMETENAWLQARQENELKRIKLDYEIVPCDKQILIEREQIITERNSTQIGNKK---D 732
QY 790 MEKMSAVQGVPRHRRGEIWFLEAQFHLKHQFP--SKQPK-DVPYKELLKQLTSQ 846
Db 733 PKVLGHAIKTVPRSKRGDWTFLAEQ-HSMNTAPVDTKRFNFNTPYHMLKLHLTEHQH 791
QY 847 AILIDGRTPPHYFSAAGAGLSYLNILKAYSLLDQEVGYCOGLSPVAGILLHNS 906
Db 792 AIFIDIGRTPFHQFYKDPGLGQSLSLNLLKAYSLDPELGYCOGLPGICGVLLHCD 851
QY 907 EEAFLKFLMFDMLGRKQYRPMITLOIQYLSRLHLDYHRLDYNHLEHEIGPSLYA 966
Db 852 ANSFQKLHLMFRNRNTKYLFDPKKFKQLYLSRLKADHLPDLVWLDQNDVSPFLYA 911
QY 967 APWLTMFASQPLGPFVAVRFDMIFLQGTVEIFKVALSLGSHKPLILOHENLETIVDI 1026
Db 912 APWLTVFSSQPLGPFVAVRFDLFLLESSDVIFKFAIALLSVHKQLLAKDNFEIMDY 971
QY 1027 KSTPLNLGLVQEKTIQNVFEMDIAKQAYEVFHVLOEELIDSPSDNORMKLL--E 1084
Db 972 KTVVPMKMEHTCMEQIMKLVFSDMIGKQLAEYKVEYNVLOEEI-----TTNHLEMLNRE 1026
QY 1085 KTNSSLRKONLDLEOLOVANGRIQSLKATIEKLLSSSEKLAQAMLTLEERSALLQTV 1144
Db 1027 KT-----QNHLEOLOQAQSSIAQLETT-----RSSQAQITTLQSQVOSLELTIQ 1073
```

```
QY 1145 ELRRSAEPSDREPECTOP 1163
Db 1074 TLGRYVQLVEHNPDLPL 1092

RESULT 8
QYTY11
ID Q9TY11 PRELIMINARY; PRT: 1379 AA.
AC Q9TY11;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE POLLUX PROTEIN.
GN PLX OR POLLUX OR CG1093.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINAL;
RA MEDLINE=99030403; PubMed=9813038;
RA Xu X.Z.S., Wes P.D., Chen H., Li H.S., Yu M., Morgan S., Liu Y.,
RA Montell C.;
RT "Retinal targets for calmodulin include proteins implicated in
RT synaptic transmission."
RL J. Biol. Chem. 273:31297-31297(1998).
DR EMBL; Y17919; CAA76939.1; -.
DR FlyBase; FBgn0004879; plx.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC. 1.
DR SMART; SM00462; TBC; 1.
DR SMART; SM00164; TBC; 1.
SQ SEQUENCE 1379 AA; 154276 MW; F9F2074CDF014BBE CRC64;

Query Match 24.2%; Score 1462.5; DB 5; Length 1379;
Best Local Similarity 32.1%; Pred. No. 6.2e-88;
Matches 366; Conservative 174; Mismatches 322; Indels 277; Gaps 30;

QY 156 QELHCPSEFDDTFKKFVLCGRRVYVAAHKAPALIDECIEKFN----- 201
Db 100 QNLSILLADISPNTHTFFEVYVGVKIRVSKRPVNTFIDDLALPKFAYDAQRLLQNRK 159
QY 202 -----HVSGRSGS 209
Db 160 MSLSEGGVIEAKPSSLSKSHDLKEEDEEQHKGHDSDSOAKPLVOLQLTGAEEG 219
QY 210 ESPRPNPPHAAPTGSOEPPVRRPMRKSFQGLRSLAFRKELQDGGLLRSSGFFSFEESDI 269
Db 220 AAPRPLEDNKE---NKSPEKPLLRGQSQIELG---HKEHSDGSPSA----- 262
QY 270 ENHLISGHNIV--OPTD----- 284
Db 263 -NSOLEAPNVIVNQPTPPRDOGVGTGTASAGPSQLHPNTAMDNIPQRDSASQGI 321
QY 285 ----IEENRTMLFTIGOSEVYLISPDTKKIALEKNEKEISFCQGIHVDFHFGICRESSG 341
Db 322 PPVQNRMTWFLVGRCDLRLSPDRKQVLLYKDFKDVASCVHGKSLDHFGIICRELNN 381
QY 342 GGGFHFCVVCFOCTNEALVDEIMMTLKQAFVAAVQQTAKAPAQL--CEGCPLOSLLHKLC 399
Db 382 DG--YIGVFKQSEHVCDIVAATAQAFTCA-EQKKQDTQIFSCHECHPMLWYHKLC 437
QY 400 ERIEGNSKTKLEQKHITTTNOEQATIFEVOKLRPRNE--ORENELIISFLRCLYE 457
Db 438 TDVEGLSEKKTQALIRRIETLSDDQEITVWAKFCGSEKTSNPVAFQNMFLMLLRAHCE 497
QY 458 EKQEKHIHIGEMKOTSOMAAENIGSELPPSATRFLDMLKNAKRLSTESLSISLRGNK 517
Db 498 SRQORHVH-----DTAENRSEFLNQYLGST-----IFMKAKRSLTNSFNLLKR----- 542
```

```

QY 518 ARGLOEHSIVDLSSTLSNTSKPSVCEKEALPISESSFKLL--GSSEDLSSDSSES 575
Db 543 -----KPS---KDDIAPVSHNLRDIRREGSAEPLGTQ--- 570
QY 576 HLPPEAPLSPQOAFRRANTLSHPPIECQEPQARGSPGVRSQK--LMRVHSYSTETP 633
Db 571 -----SPGFRSRNTVGSF-----SKPTAEQLKSPMDIFIKVGNP 611
QY 634 HERKDFESKANHLGDSGGTPVTRHRSRQOIFLRVATPQKACDSSSYEDYSELGELPP 693
Db 612 KE-----AETHGSRWQAILNSVWTPSKGLDSEVPTEFLS----- 646
QY 694 RSPLEPCVCEDPGPPPEEKRTSRELRLAQKALIQILLRMEKENQKLOASENDLLN 753
Db 647 -----PMRPAKRGKRDAAELRELMTAIRQIMLRNMTENAMLOARQENEL 695
QY 754 KRLKLDYEEITPCLEVTVMKML---STP-GRSRIKFDMEKMSAVGGVPRHRRGEI 809
Db 696 KRIKLDYEEIVPCDQLIERNEQIERNSTQIGNKK---DPKVLGHAIRTVGPRSKRGDV 752
QY 810 WKFLAEQHLKHQFP--SKQPK-DVPYKELLKQLTSQOHAIDLGRTPPHYFSAQL 866
Db 753 WTFLEAQ-HSMNTAPVDTRKFPNFTPYHMLLKLHTEHQHAIFIDLGRTPPHYFSAQL 811
QY 867 GAGQLSLYNILKAYSLLDQEVGYCOGLSFVAGILLHMESEEAFAKMLFDMGLRQY 926
Db 812 GUGQLSLFNLLKAYSILDPGLGYCOGLFCIGVLLHLCDEANSFOLLKHLFRNRMTKY 871
QY 927 RPDMLIIOQYLSRLHLDYHRLYNHLEHEIGPSLYAAPWFLTMFASQFPLGFVARV 986
Db 872 LPDMKKFQOLYQLSRLVKDHLPLDLYVWLDQNDVSPTLYAAPWFLTMFASQFPLGFVARV 931
QY 987 FDMFLOQTEVIFKVALSLLGSHKPLILOHENLETVDFIKSTLPNLGLVQMEKTINQVF 1046
Db 932 FDLLESSDVFIFKFAIALLSVHKQQLAKDNFEEIMDYLTVPKMEHTCMEQIMKLVE 991
QY 1047 ENDIKOLQAYEVEVHVLQELIDSSPLSDNORMDKL--EKTNSSLRKQNLDLLEQLQVA 1104
Db 992 SMDICKQLAEVNEVYNVQEEI-----TTNHLEMLNREKT-----QNHLEQLOLFA 1040
QY 1105 NGRIOSEATIEKLLSSSKLQAMLTLELRSALLQTVLEELRRRSAPSDREPECTOP 1163
Db 1041 QSSIAQLETT-----RSSQQAQITTLQSQVQSLELTQTLGRYVQGVLEHNPDLPL 1092

RESULT 9
ID Q24193 PRELIMINARY; PRT; 732 AA.
AC Q24193;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE POLUX.
GN PLX OR CG1093.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OREGON R; TISSUE=TRACHEA, CNS;
RA Zhang S.D., Kassisi J., Olde B., Mellerick D.M., Odenwald W.F.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50542; AAB02200.1; -
DR FlyBase; FBgn0004879; plx.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS01179; PID; 1.
SQ SEQUENCE 732 AA; 83338 MW; 02707F4591775283 CRC64;

```

```

Query Match 16.6%; Score 1007.5; DB 5; Length 732;
Best Local Similarity 46.8%; Pred. No. 3.1e-58;
Matches 217; Conservative 79; Mismatches 137; Indels 31; Gaps 10;

QY 709 PPEEKRTSRELRLQKAILQOILLRMEKENQKLOASENDLLNKRLLKLDYEEITPCIK 768
Db 4 PAKRKRDAAELRELMTAIRQITIMLRNMTENAMLOARQENELKRIKLDYEEIVPCDK 63
QY 769 EVTTVMKML---STP-GRSRIKFDMEKMSAVGGVPRHRRGEIWKFLAEQHLKHQFP 824
Db 64 OLIEREQIERNSTQIGNKK---DPKVLGHAIRTVGPRSKRGDVWTFLEAQ-HSMNTAP 119
QY 825 --SKQPK-DVPYKELLKQLTSQOHAIDLGRTPPHYFSAQLGAGQLSLYNILKAYS 881
Db 120 VDKRFPNFTPYHMLLKLHTEHQHAIFIDLGRTPPHYFSAQLGAGQLSLYNILKAYS 179
QY 882 LLDQEVGYCOGLSFVAGILLHMESEEAFAKMLFDMGLRQYRPMIILQIQMYQLS 941
Db 180 ILDPGLGYCOGLFCIGVLLHLCDEANSFOLLKHLFRNRMTKYLPDMKKFQOLYQLS 239
QY 942 RLLHLDYHRLYNHLEHEIGPSLYAAPWFLTMFASQFPLGFVARVDMIFLOGTEVIFKV 1001
Db 240 RLVKDHLPLDLYVWLDQNDVSPTLYAAPWFLTMFASQFPLGFVARVDFLLESSDVFIFK 299
QY 1002 ALSLLGSHKPLILOHENLETVDFIKSTLPNLGLVQMEKTINQVEMDIKOLQAYEVEY 1061
Db 300 AIALSVHKQQLAKDNFEEIMDYLTVPKMEHTCMEQIMKLVSMDIGKQLAEVNEVY 359
QY 1062 HVLQELIDSSPLSDNORMDKL--EKTNSSLRKQNLDLLEQLQVANGRIQSLAEATEIKLL 1119
Db 360 NVLQEEI-----TTNHLEMLNREKT-----QNHLEQLOLQFASSIAQLETT----- 403
QY 1120 SSKLQAMLTLELRSALLQTVLEELRRRSAPSDREPECTOP 1163
Db 404 --RSSQQAQITTLQSQVQSLELTQTLGRYVQGVLEHNPDLPL 445

RESULT 10
QY3P9
ID QY3P9 PRELIMINARY; PRT; 997 AA.
AC QY3P9;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE RAB6 GTPASE ACTIVATING PROTEIN, GAPCEN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=99219856; PubMed=10202141;
RA Cui M.H., Possmayer F., Zander H., Bordes N., Jollivet F.,
RA Couedel-Courteille A., Janoueix-Lerosey I., Langsley G., Bornens M.,
RA Goud B.;
RT "Characterization of GAPCEN, a GTPase activating protein for Rab6,
RT part of which associates with the centrosome.";
RL EMBL J. 18:1772-1782(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Cui M.H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011679; CAB40267.2; -
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00640; PID; 1.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS01179; PID; 1.
SQ SEQUENCE 997 AA; 114142 MW; E3694B6FA9539C52 CRC64;

```

Query Match 9.6%; Score 580; DB 4; Length 997;
Best Local Similarity 23.9%; Pred. No. 9e-30;
Matches 221; Conservative 137; Mismatches 311; Indels 254; Gaps 33;

```
QY 295 IGSEVLLSPDKKIALEKNFKIEISFCOGIRHDFHG-----FICRESSGGGFGHFV 348
DB 119 VSGIWRLLDPQNTTEIANPIYKILFCVNG-----HGTGPSDCFAFTESHYNABEL-FR 172
QY 349 CVYFOCTNEALVDEIMMILKQAFVAAVQOTAKAPQAQCEGCPQLSHLKCERIEGMNNS 408
DB 173 IHVERCEIQEAVSRILYSFATAPRRSAKQTPLSATA--APQTPDSDFTFVSVLE----- 225
QY 409 KTKLEQLQKHTTTNQBQATIFEVOKLPRNQENELIISPLRC--LYEEKQKEHHI 466
DB 226 -----IKEDDGKGYFSAV-----PKDKRO-----CFKLROGDKKIVIV 260
QY 467 GEMKQTSMAAENIGSELPPSAFRFLDKLKNKAKRSLSLESILSRGNKARGLQEHSI 526
DB 261 VQQTNNKELAIERC-----FGL-----LLSPGKDVNSDMHLL 293
QY 527 SVLDLSLSLTSLNTSKPSVCKEALPISSESKLLGSEDSSESHEL-----PEE 580
DB 294 --DLES-----MGK-----SSDGKSVITGWNPKS 317
QY 581 PAPLSQQAARRRANTLSHPPIECQBPQPARGSPGVSRQKLMRYHSVSTETPHERKDFE 640
DB 318 P-----HFQV-----VNEETPKDKVLFM 335
QY 641 SKANHLGDSGGTPVKTRHRSWRQOIFLRVATPQKACDSSSRVEDYSELGELPPRSPLEPV 700
DB 336 TTAVDL-----VIT-----EVQSPVRFLLETKVRVCSNPERL----- 367
QY 701 CEDGPGPPPEEKRRSRE-----LRELWOKA-----ILQQILLRMKEKNKLO----- 745
DB 368 -----FWP-----FSKRSTTENFFLKQIKQIKRERKNNTDTLYEVVCLSESESERRRKTAS 419
QY 746 -----ASENDLLNKLRLDYEE-----ITPCLKEVTVTWKMLSTPGRS 784
DB 420 PSVRLPQSGSQSVIISPDPDEEDENDEPILLSGSDGVSKCAEKILETWGELLS-KWHL 478
QY 785 KIKEDMEKMSAVGQVPRHHRGEIMKFLAEQFHLKHQFPKQOPKDVPIYKELLKQTSQ 844
DB 479 NLNVRPKQLSSLVNGVPEALRGVEVQWLLAGCHNNDHLVEK-----YRIILITRESQP 530
QY 845 OHAILDLGRTFTTPHYFSQAQAGQLSYNLKAYSLLDQEVGYCOGLSFVAGIILLHM 904
DB 531 DSAITRDINRTFPAHDYFKDTGQDGDLSYKICKAYSVDYEIGYCOGQSFLLAAILLHM 590
QY 905 SEEEAFKMLKFLMDFMGLKQYRPMIILQIQMYQLSRLLHDYHRLYNHLEHEIGPSL 964
DB 591 PESQAFSLVKIMFDYGLRELKONFEDLHCKEYQLERLMQOEYIPOLYNHFLDISLEAHM 650
QY 965 YAAFWELTFPASOPPLGFFVARVDFMIFLAGTEYIFKVALSLGSHKPLIIQHENLETIVD 1024
DB 651 YASOWFLTLTAKFPLMYVPHIIDLCEGISVFNVALGLLATSXDLL--TDFEGALK 709
QY 1025 FIKSTLPNLGLVOMEXTINQVFE-----MDIA-KOLQAYEYVHVLOBELI--DSSPLSDNQ 1078
DB 710 FFRVQLPK--RYRSEENAKKLMELACNMKISQKKKYEYHTMREQAQQEDPIERFE 767
QY 1079 RMD-KLEKTNSLRKNLDLEOVANGRIQSIETATIEKLLSSESKLKQAMLTLELERS 1137
DB 768 RERNRQOENNRLEQNDLHAHEL-----VTSIALRKLONAEKADALNKELLMTQK 821
QY 1138 ALLQTVBELRRRSRSPDREPEC 1160
DB 822 KLIDAEERKRLSEESAQLKEMC 844
```

RESULT 11
Q967F5

```
ID Q967F5 PRELIMINARY; PRT; 1136 AA.
AC Q967F5:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F35H12.2B PROTEIN.
GN F35H12.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Nelson J., Gattung S.;
RT "The sequence of C. elegans cosmid F35H12.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41540; AAK39232.1; -.
SQ SEQUENCE 1136 AA; 131257 MW; A86F4E3448103AE7 CRC64;
```

Query Match 8.6%; Score 523; DB 5; Length 1136;
Best Local Similarity 28.7%; Pred. No. 6.4e-26;
Matches 155; Conservative 80; Mismatches 181; Indels 124; Gaps 18;

```
QY 660 SNRQOIFLAVATPOKACD-----SSRYEDYSELGELPPRSP---LEPV 700
DB 347 AWQLEGVYFLTPPSKSDQSDPNDRKLTFTISLSDSRKSKONLGSFSPMTQLLHPT 406
QY 701 -----CEDGPGPPPEEKRTSRELWOKAILQOILLRMKEKNQKQASENDLLNK 754
DB 407 GDDESDCDEPILLSGSKVQCEKEHEHMDQLI-----EN----- 442
QY 755 RLKLDYEEITPCLKEVTVTWKMLSTPGRSKIKFDMKMSAVGQVPRHHRGEIMKFLA 814
DB 443 -----WDQQSDRP-----QKISELYLDGIPDKLGRVWOLLS 474
QY 815 EQFHLKHQFPKQOPKDV-PYKELLKQLTSQQHAILDLGRFTPTPTHTFSAQLGAGQLSL 873
DB 475 NAI-----DQDLVKEHYHFLSQCPSEQVIMRDIHRTFPANDYKESQKQGOQL 525
QY 874 YNLIKAYSLLDQEVGYCOGLSFVAGIILLHMSSEAFKMLKFLMDFMGLKQYRPMIIL 933
DB 526 YKISKVYSLVDEVSVCQGLSFLAASLLHMPPEQAFCTLVKIMFNKGLRDLFKLGFNDL 585
QY 934 QIQMYQLSRLLHDYHRLYNHLEHEIGPSLYAAPFLTFASFQPLGFFVARVDFMIFLQ 993
DB 586 HLRFFQLTALLKDYIPDLSSHLEHIGIETHMYASQWFLTFTAKFFLQWVFFFLDILFSLQ 645
QY 994 GTBVEIKVALSLGSHKPLIIQHENLETIVDFIKTSLPNLGLVQ-----MEKTIQNVFE 1047
DB 646 GMTTIFHISALLDDAKTDLQLD-FEGTLKYFRVSLPRKYRFEASTKCLIHAKV--FR 702
QY 1048 MDIAKOLQAYEYVHVLOE-ELIDSSPLSDNQMDK-----LEKTNSSLRKKQ-- 1093
DB 703 LNHSK-LEVYENYKRIKELERENEDPV---LRMEKEIGHQANTLRLENDDLAHEL 758
QY 1094 --NLDLLEQLOVANGRIQSIETATIEKLLSSESKLKQAMLTLELERSAL--LQVVELRR 1149
DB 759 TSKIELRRKLDVAEDQIETSAIERL-----TRQNMIDILEENKNLMREYEQIKEMYRR 812
```


AC	Q96CM4;
AD	01-DEC-2001 (TReMBLrel. 19, Created)
AE	DT
AF	01-DEC-2001 (TReMBLrel. 19, Last sequence update)
AG	DT
AH	01-DEC-2001 (TReMBLrel. 19, Last annotation update)
AI	DE
AJ	STIMILAR TO ECOTROPIC VIRAL INTEGRATION SITE 5.
AK	OS
AL	OS Homo sapiens (Human).
AM	OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AN	OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AO	OX NCBI_TaxID=9606;
AP	[1]
AQ	SEQUENCE FROM N.A.
AR	RP TISSUE=PLACENTA, AND CHORIOCARCINOMA;
AS	RA Straussberg R.;
AT	RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
AV	DR EMBL; BC014111; RAH14111.1; -
AW	SQ SEQUENCE 794 AA; 91375 MW; 598B06DD2F098664 CRC64;
AX	SD

```

Query Match      8.3%; Score 500.5; DB 4; Length 794;
Best Local Similarity 26.4%; Pred. No. 1.2e-24;
Matches 146; Conservative 106; Mismatches 198; Indels 103; Gaps
QY 669 VAPQKACDSSRRYEDYSELGPPRSPLEVPCEGPPGPPPEKKRTSRRLRWOKAI 728
      :|:|: ||||: | | | | | | | | | | | | | | | | | | | | |
Db 1 MASPTLSPDSSQ-----EALSAPTC-----SPTSDSENLSPOELE----- 36
QY 729 LQOILLRLMEKNKIQLOASENDL-----LNKRKLQDYEEITPCIKR 769

```

DD 770 VTTWEKMLSTPGRSKIFEDMEKNHSAVGGQVPRRHRGETWKFLAEQFHUKHQFPSPKQP 829

3/ - LIAKLEEQNKULEADSKMSKMSWNGSKNSGSSLVSSSSASASNSUSHEBEDIWLNUR 34

[illegible]

QY	1010	KPILQHENLETIVDFIRSTLPNLGLVQMEXTINQVFFEMI-AKQLQAYEYEVHVLQ---	1065
DB	316	QAEMLQLD-MEGMSYQFQRVIPHOEDSPDKLVUKAYOVKNPKMKRLEKEYAAMAKSKE	374

```

QY      ELIDUSSFV - SUN --- QWUKLEK  INSURKQRNMLDUEUUYVATNAGLQJGUAH :
|| :: || :: || :: || :: || :: || :: || :: || :: || :: || :: || ::
375 MEQIIEIKRLRTENRLLKORIELEKGVTQAQAEENYVIKRELAVVRQCSCSAEDLQ 434
Db

QV      KLLSSRSKKUO-----AMLTLELERSALLOT-----VEELRRRSA 1151

```

```

DB      435 KAQSTLKLQEQENPKRTIEFVSHLETLSQSKUKETILGALREKQADVYDREARJCS 434
QY      1152 EPSDREPECTQPE 1164
      | : | |
Db      495 LPDENNVAAIOLFF 507

```

Search completed: August 28, 2002, 16:28:34
Job time: 322 sec

RESULT	14
O60447	
ID	PRELIMINARY;
AC	PRT; 810 AA.
DT	O60447;
DT	01-AUG-1998 (T-EMBLrel..07, Created)
DT	01-AUG-1998 (T-EMBLrel..07, Last sequence update)
DT	01-DEC-2001 (T-EMBLrel..19, Last annotation update)
DE	EVI-5 HOMOLOG.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBTaxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-BRAIN, IVER.
RA	Roberts T.P., Cowell J.C.;
RT	"Human EVI5 gene complete cdna sequence.";
RL	Hum. Mol. Genet. 0:0-0(1998).
DR	EBML; AF008915; AAC16031.1; -.
DR	InterPro; IPRO00515; BPD.transp.
DR	InterPro; IPRO00195; RabGAP_TBC.
DR	pfam; PF00566; TBC; 1.
DR	SMART; SMO0164; TBC; 1.
DR	PROSITE; PS00402; BPD_TRANSP_INN_MMEMBR; UNKNOWN_1.
SQ	SEQUENCE 810 AA; 92847 MW; 8988011991329BF9 CRC64;

Query Match	8.4%	Score 510.5;	DB 4;	Length 810;
Best Local Similarity	28.5%;	Pred. No. 2.6e-25;		
Matches 149;	Conservative	99;	Mismatches 171;	Indels 103;
Gaps	17;			

QY		684	QIFRWATPQKADSSRRSIEDISLGBSPFSPLEPCEDGPPFPPEEKKRRTSKRELKL	723
Db	:	44	OMASQVASTPSLHTSSSTLS-----TPALSSESP-----SQLSPDVLEL	85
QY	:	724	WQKALQQIILLRMKEKNKLOASENDLLNKRKL-----DYEITPCL-KEVT	771
Db	:	86	LAKLEQNILL---ETDSKLSRYNGSRNRSGSLSSASSNSHLSEDSWTLGWRIV	142
QY	:	772	TVMKMLSTPGRSKIPTDKMEKMHSAGVGQVPRIHRHGEINWKFLABQPHLKHQFPSKQPKD	831

QY 832 VPKELLKQTSQOHAII DLGRTFPTHFVSQLGAGLSLYNILKAYSLLDQEVGYCQ 891
 188 --YSELLKWTSPCEKLIRRDIAITYPEHNFKEKDSLGQEVLFVWMKAYSILNDQEVGYCQ 245
 Db

Db	246	GS	A	F	I	V	G	L	L	M	P	E	S	E	A	F	C	V	F	K	L	M	O	D	Y	R	L	R	E	L	F	K	S	M	A	E	L	G	I	C	M	Y	Q	F	E	C	M	I	O	E	H	L	P	E	L	305
QY	952	Y	N	L	E	E	H	E	T	G	P	S	I	Y	A	P	W	F	T	M	F	A	S	O	P	P	L	G	T	E	V	F	K	V	A	L	S	L	L	G	S	H	K	P	1011											

QY	1012	LILQHENLETIVDFIKSTLPNLGLVQ-MEKTINQVFEMDIKOLQ-AYEVEYHVLQEELI	1069
		::: :: : : : : :	
Dd	366	ELMQ-----LDMEGMLHQFVKVPHQFDGPPDKLIQAAYQVY-----	403

DB 404 -----NSKKMKALEKEYTTIKTAEME-----EQVE-----IRLRRTENLLRQRI 443

QY 1130 LTLELERSA-----LLQTVLEELRRRSAPSDREPECTOPE 1164

RESULT 15
Q96CN4
ID Q96CN4 PRELIMINARY;
PRT; 794 AA.

RESULT 15
Q96CN4
ID Q96CN4

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:23:52 ; Search time 15.15 seconds
(without alignments)
2985.111 Million cell updates/sec

Title: US-09-762-311-5
Perfect score: 6055
Sequence: 1 MPEIIFARKHLLPNEVSD.....RSAEPSDREPTQPEPTGD 1168

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	387	6.4	YD99_SCHPO	Q10496 schizosacch
2	343	5.7	YAD4_SCHPO	Q09830 schizosacch
3	317.5	5.2	YM52_YEAST	Q04322 saccharomyc
4	274	4.5	MIC1_YEAST	P53258 saccharomyc
5	247.5	4.1	YN33_YEAST	P48566 saccharomyc
6	232.5	3.8	CENF_HUMAN	P49454 homo sapien
7	227	3.7	MYHB_MOUSE	O08638 mus musculu
8	217	3.6	MYHB_HUMAN	P35749 homo sapien
9	212.5	3.5	MYHB_CHICK	P10587 gallus gall
10	208	3.4	GOG4_HUMAN	Q13439 homo sapien
11	207	3.4	CENE_HUMAN	Q02224 homo sapien
12	202	3.3	MYHB_RABIT	P35748 oryctolagus
13	197	3.3	MYHA_BOVIN	Q27991 bos taurus
14	196.5	3.2	TMF1_HUMAN	P82094 homo sapien
15	195	3.2	MYH9_HUMAN	P35579 homo sapien
16	195	3.2	MYHA_HUMAN	P35580 homo sapien
17	195	3.2	MYSJ_SCHPO	O14157 schizosacch
18	193	3.2	MYSD_DICDI	P54697 dictyostelli
19	193.5	3.2	MYH2_HUMAN	Q9ukx2 homo sapien
20	191.5	3.2	ACF7_HUMAN	Q9upn3 homo sapien
21	190.5	3.1	TPR_HUMAN	P12270 homo sapien
22	189	3.1	ALM1_SCHPO	Q9utk5 schizosacch
23	189	3.1	MYHA_RAT	Q9jlto rattus norv
24	187.5	3.1	YD86_SCHPO	Q10411 schizosacch
25	187.5	3.1	MYSD_DICDI	P08799 dictyostelli
26	186.5	3.1	HMW2_MYCSE	P47460 mycoplasma
27	186.5	3.1	MYH1_HUMAN	P12882 homo sapien
28	186.5	3.1	DMD_CANFA	O97592 canis fami
29	186.5	3.1	AKA9_HUMAN	Q99996 h a-kinase
30	184.5	3.0	MYH9_RAT	Q62812 rattus norv
31	184	3.0	MYH7_PIG	P79293 sus scrofa
32	182.5	3.0	MYSN_DRONE	Q98323 drosophila
33	182	3.0	MYH6_MESAU	P13539 mesocricetu

Query Match 6.4%; Score 387; DB 1; Length 1031;
Best Local Similarity 20.7%; Pred. No. 5.4e-13;
Matches 207; Conservative 153; Mismatches 377; Indels 262; Gaps 39;

34 181.5 3.0 1940 1 MYH3_CHICK P02565 gallus gall
35 181 3.0 1186 1 XPG_HUMAN P28715 homo sapien
36 181 3.0 1325 1 G160_MOUSE P55937 mus musculu
37 180.5 3.0 472 1 YAE3_SCHPO Q09844 schizosacch
38 180.5 3.0 1818 1 HMW2_MYCPN P75471 mycoplasma
39 180.5 3.0 1935 1 MYSS_CYPCA Q00339 cyprinus ca
40 180.5 3.0 1937 1 MYH8_HUMAN P13535 homo sapien
41 180 3.0 1934 1 MYH7_MESAU Q02566 mus musculu
42 180 3.0 1938 1 MYH6_MOUSE P02563 rattus norv
43 180 3.0 1938 1 MYH6_RAT P02562 oryctolagus
44 179 3.0 1084 1 MYSS_RABIT P29616 gallus gall
45 177 2.9 1102 1 MYSC_CHICK

ALIGNMENTS

RESULT 1
YD99_SCHPO
ID YD99_SCHPO STANDARD; PRT; 1031 AA.
AC Q10496;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 116.3 kDa protein C26F1.09 in chromosome I.
GN SPAC26F1.09.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: SOME, TO YEAST YMR192W AND YPL249C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73100; CAA97366.1; -
DR InterPro: IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 708 818
FT CARBOHYD 113 113 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 438 438 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1031 AA; 116289 MW; F2579A4C21D79EC CRC64;

QY 206 SRGSESPRPNPHAAPTGSGQVRRPMKRSFSOPGL-RSLAFRKELODGLRSSGFFSF 264
 Db 144 SESETPVNEVNETATLGNEDVGE---RSGFPSEGLDNEPESQORDLDG-----NL 192
 QY 265 EESDIENHLISGNIVQPTDIENRTMLFTIQSEVYLISPTKKIALE-----KNFKEI 319
 Db 193 APEDLKDEKVSHEFNEPDLQO-----QEEYSDDDTNVNEFEDVNEIENEHQL 243
 QY 320 SEC--SOGIRHV-DHGFCTCRSSGGGPHFYVFCQCTNEALVDIEMTLKQAFVAAV 376
 Db 244 SVADEQTSRLVKGKIMFVKGEDFGEA-----DISNSVFIEQ---NGPNSDVTSGF 292
 QY 377 QOTAKAPALCEGCPQLQSLHKLCERTEGNSMKTLELQKHLTLTINQOQATIFEVOKL 436
 Db 293 KETSSI-----VNSST---TEKPGVALDSQNDTIF-----321
 QY 437 RPNREORENELIISFURLYEKEKQHHIGEMKQTSQMAAENIGSELPPSATRRFLDML 496
 Db 322 ---NEQTNLSLETF-----NDLTLDHL-----PENVESE--PVAGK-----353
 QY 497 KNAKRSLTESLESILSRGNKARGLQEHHSISYDLSDSLSTLSNTSKPSVCEKE-----551
 Db 354 -----ENETAKNESGASNDHKNVHVFLKSEDAITLNEEKIATQDDPLEA 401
 QY 552 ALPISESSFKLLGSDSDSHLPEEPAPLSPQAFRRRANTLSHPPIEC-----OE 606
 Db 402 PPIVASSSTIFLNSNORDELASGSGQEPHP-----KGTNSTSSPLDTNNLSNE 454
 QY 607 PQPARGSPGVQSKMLRVHSVSTEPHERKDFESKANHLGDSGGTP-----VKTRRHWS 661
 Db 455 PPSHVLDASSETIEVQTIKLQNOVPETIKDEVGKN-TAFSPGTSLSNTHVTKRSRA 513
 QY 662 RQOIFLRVATPOKACDSSSKRYEDYSELGLPPR---SPLEPVCEDEGPFPPPEEKRTSR 718
 Db 514 HNSTSPFTAVSSWLNPLRYP-----SKSPRVTSYLSYFISKP-----R 556
 QY 719 ELRELQKAILQOILLLRMEKENQKLOASENDLLNKLKLYEETPCLEKVE-----770
 Db 557 SIGDAQKLEILEVLOSQSTVSNOVFTLLSNFTQNPFLVD-----ECDFENRLTMHN 611
 QY 771 -----TTVKEKMLSTPGRSKIFDMKHS-----AVGQGVPRHHRGEIWKFLA 814
 Db 612 SHTVHTVVMKTI-----SWTSYDYEMQYSSISIKNCDSKAIKDLDTFAPEI-----661
 QY 815 EQFHLKHQFPSSKQ--PKOVVPYKELLQKLTQSOHAILDLGRTPFPHYFSAQLGAGQLS 872
 Db 662 ---LSHFFSNRQLEPTD-----NIAESTAN 684
 QY 873 LYNILKAYSLLDQEVGYCGGLSFVAGILLHMSSEAFKMLKFLMDFMGLRKOYRPMII 932
 Db 685 LHRVLSLAIVLPQVGYTCQMSWAGALLMLHPAQOAFALLVFLKNYHLQNIIFSSEMRG 744
 QY 933 LQIQMYQLSRLLDYHRDLYNLEERHEIGPSLYAAPWELTMEASOPPLGFVARVDFMIFL 992
 Db 745 LSRVHQFTRLVEDYMPSLAIFKRODITKCTSYASEWFLTLFAYKPLEVVAHLVDILFL 804
 QY 993 QGTEVIFKVALSLGSHKPLILQHENLEIVDFIKSTLENLGLVQMEKT-----1041
 Db 805 YGPGILFNGGLALL-SHSQESLLKLNMDRLISYLKEDI---FLAPKEQEGENYDTSLF 859
 QY 1042 INOVFMDIAKO-LQAYEYHYVL--OEBLIDSSPLSDNORMDKLEKTNSSLRKONL-----1095
 Db 860 VKTAFSEIQPOVLDRYGNEYDILLKSEHLDSS-----LEEMNRHKSLSNEHFIMLS 912
 QY 1096 DLLEQLOVANGRIQSLEATIEKLLSSEKLNQAMLTLEL 1134
 Db 913 DSMANLQVEHENMSAL-LLKEKMYLNQOTVEQASLKSEI 950

RESULT 2
 YADA_SCHPO
 ID YADA_SCHPO
 AC Q09830;

PRT: 772 AA.

DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 86.2 kDa protein C4G8.04 in chromosome I.
 GN SPAC4G8.04.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Badcock K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO YEAST MIP1 AND YNL293W.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z56276; CAA91205.1; -
 CC InterPro: IPR000195; RABGAP_TBC.
 CC Pfam: PF00566; TBC; 1.
 CC SMART: SM00164; TBC; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 772 AA; 86232 MW; BADCLC8B8EB527E7 CRC64;

Query Match 5.7%; Score 343; DB 1; Length 772;
 Best Local Similarity 22.1%; Pred No. 7.4e-11;
 Matches 198; Conservative 126; Mismatches 340; Indels 232; Gaps 34;
 QY 294 TIQSEVYLISPTKKIALEKFNKEISFCQGIHVDFHGFCTCRSSGGGPHFYVFCQ 353
 Db 4 SVSDSSVYLLSSATKPIA-----SGSVATG-----28
 QY 354 CTNEALVDIEMTLKQAFVAAVQOTAKAPALCEGCPQLQSLHKLCERTEGNSMKTLE 413
 Db 29 -----IDS--MPSKVDTIIPCDDLLENSKSAFLFVECNQESLHSPGSLHLPVDSIEL 80
 QY 414 LOKH-LTTLTNEQATIFEE---VQKLRPNQRENELIISFLRCLYBEKQEH-----IHI 466
 Db 81 IEKHGAVNLLRLQALDAERDSFISDLKPFESRE-----YVFRELLREHGLDPLV 131
 QY 467 GEMKQTSQMAAENIGSELPPSATR-----FRDMLKNK-----AKRSLTESLE 509
 Db 132 ANTKLSQRHSASFFPSQSPSPSPENPSSLTGKPHLYARIDSAINEPFTPSDRLSPSLV 191
 QY 510 STLSRGNKARGLQEHHSISYDLSDSLSTLSNTSKPSVCEK-EALPIS-----SS 559
 Db 192 PLL-----KLPAI-DHVAVSSSSSDLPDPNSASYIASSKQASSLKLTSSLKFFYWTSS 246
 QY 560 FKLLGSSELDSDSESHLPEEPAPLSPQAFRRRANTLSHFPICEQPPQARGSPGVQ 619
 Db 247 SLSQHTRENHLNHDSTSLRDHDPILLSKSKFFRSP-----RCSTPVSSTFFVSATSE 298
 QY 620 RKLMYHVSSTPTHERKDFESKANHLGDSGG-----551
 Db 299 PEVEY-SVSTKSSSNKLNRLSLSKLLSTSLNKNKPLSLSTAPSMPISIGSVELGNMIP 357
 QY 652 --TPVKTRRHSRQOIFLRVATPOKACDSSRYEDYSELGLPPR-SPLEPVCEDEGPF 708
 Db 358 KETQPPSMRNDWKDYL-----DNNSK-EILDQFGFLQKRPSTHTPLC-----398
 QY 709 PPEE-----KKRT-----SRELRELQKAIL---QOILLRMEKENQKLOA 746
 Db 399 -PEDIKLNKQYLSFYESNYGLVDFGNEGLDGLNDSPLLNNKDLILLDMKESTROKNWS 457

```

QY 747 SENDLLNKRLLDYETTPCLKEVTWVEKMLSTPGRSKIKFDMKMSHSAVQGVPRHHR 806
Db 458 LFQRLYKXKIT-DEDTGLIGTSSIGV-----GRHG-KKRWKHFRELKNGVPLCYK 510
QY 807 GEIKWFLAEQFHLKHQFPKQPKDVP---YKELLKQLTSQOHA-----ILIDGRFTFPH 860
Db 511 AKVMECSGAQL-HS-----PGYYELLRSRTDEVESASVAQIDMINTMAKNV 559
QY 861 YFSAGLQAGQSLYNILKAYSLLDQEVGYCGQLSPV-AGIILLHMESEAFKMLKFLMFD 919
Db 560 FFGGR-GGIPKRLRLVAYSRHNPHIGYCQGMNVIGAFLLLLYASEEDAFLMLSIEN 618
QY 920 MGLRQYRPMILLOIQMYQLSRLLHDYHRLYNHLEHEIGPSLYAAPWFLTFPASQFP 979
Db 619 VLPKRYFTPDWTSRADQLVKFSKESLPEIYSHLELGLVDLDAISPHWFLSVYTDLP 678
QY 980 LGFVARVDFMFLQGTVEIFKVALSLGSHKPLILLOHENLETVDFIKSTLPNLGLVQME 1039
Db 679 TNISFRIDMLFCDGYVCLFRVALTILKSLKQIILACNSSAIVSFLSDL----- 728
QY 1040 KTIQNVFEMDIQALQAYEVHYHVLQELIDSSPLSDNORMDKLEKTNSSLRKONL 1095
Db 729 -----VOYSEPDSEFIKEAA-----DRWSKLVTSEKSIERKNL 761

RESULT 3
YMS2_YEAST STANDARD; PRT; 720 AA.
AC Q04322;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 82.1 kDa protein in SGS1-MRPL24 intergenic region.
GN YMR192W OR YM9646.04.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YPL249C AND S.POMBE SPAC26F1.09.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 247815; CAA87813.1;
DR SGD; S0004804; YMR192W
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
KW Hypothetical protein.
SQ SEQUENCE 720 AA; 82131 MW; E1423DB4F15F7267 CRC64;

```

```

Query Match 5.28; Score 317.5; DB 1; Length 720;
Best Local Similarity 21.08; Pred. No. 1.5e-09;
Matches 149; Conservative 126; Mismatches 244; Indels 191; Gaps 26;

QY 557 ESSFKLLGSSDLGS--DSSEHLPEE-----PAPLSPQAFRRRANTL--SHFPICEQ 606
Db 92 ESQKTIPLSDLSQLETSESKVEEALKRTTSPPLP-----RADCIESAALKSSL 145
QY 607 PP-----OPAGSPGVSGVSKLMRYHSVSTETPHERKDFESKANHLGDSGGTPVKTRR 658
Db 146 PPVLAGNKDQAPLDRPOLPPQVNAETLHKAPHGN----- 183

```

```

QY 659 HSWROQIFLRVATPOKACDSSSRVEDYSELGELPPR-----SPLEPVCEDBGPGPPPEKKR 715
Db 184 -----ATPSKSPTSVAGNSSSTPTPLPRRIEDPLDLAAQ----- 219
QY 716 TSRELRELWQKAILQOILLRMEXENOKLOASENDLLNKRLLKLDYEEI--TPCLKEVTTV 773
Db 220 -----KHFLASTFRKRMFLFYKSEDSNISKCDLDKN-ILN--LKDSKKINNNEIPEVSSF 271
QY 774 WEKMLSTPGRSKIKFDMKMSHSAVQGVPRHHRGIBTWFLAEQFHLKHQFPKQPKDVP 833
Db 272 WLKVIQDYQNTLIN-DIETLHQLSRGIPAAAYRLVWQLVSV-----AKSFSFDP 321
QY 834 YKELLKQLTSQOHAAILDLGRFTFPHYFSAQLGAGQLSLY-----NLTKAYS 881
Db 322 YETVLTMAP-----FDVQEFENQLKMDMEVPSEYVVKRISNVLKAYL 363
QY 882 LLDDEVGYCGQLSFVAGILLHLMSEEE--AFKMLKFLMFDGLRKQYRPMILLOIQMYQ 939
Db 364 LFDPECEFTDIAYIIN-MILDVCEEAEANAFGLVRLMKVYGLRLLFLPLPSASEIDILCYK 422
QY 940 LSRLLHDYHRLYNHLEHEIGPSLYAAPWFLTFASQFPLGFVARVDFMFLQGTVEIF 999
Db 423 FDLRVEEYPEIHNHMYEKGVRSSMFLPGFPTTLFQKKLPTFQIPRIGDMVFLGIDSIM 482
QY 1000 KVALSLGSHKPLILQHENLETVDFIKSTLPNLGLVQMEKT-----INQVF 1046
Db 483 RILATLLSNSRDLHK-MGFDDMLELLKSLGLLDAYIKQNDCTRGDTLLSNCEMDKLQDS 541
QY 1047 EMDIA---KQLQAYEVEYHVL-----QEELIDSSPLSDNORMDKLEKTNSSLRKON 1094
Db 542 MMKVAITPKTKMKYSSEYEEIHRLDNKEVQYKSTENKHLQKHVRKLENDYTSLNREH 601
QY 1095 LDLLEQ-----LQVANG-RIOSELEATIEKLISSESK-----LKQ 1127
Db 602 VTIANELVKRLNLTIESVLNENNGYKLIQILD--LKKLDSEKKQVLGVVVPNDLKKOLEE 659
QY 1128 AM-----LTLELERSALLQTVELRRRSAPSPREPECTQPEPTG 1167
Db 660 TMKKNTQVMDENLKLQDRISLERLIEIKTANKNGTLFEYSNKNPNPLG 709

RESULT 4
MICL_YEAST STANDARD; PRT; 950 AA.
AC P53258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MIC1 protein.
GN MIC1 OR YGR100W OR G5717.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Serpe M., Kosman D.J.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hernandez K., Weber N., Wipfli P., Schmidheini T.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERACTS WITH MAC1.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO S.POMBE SPAC48.04, SPCC1259.11C, SPBC215.01 AND
CC YEAST YNL293W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U52667; AAB01977.1; -
CC SCD: 272885; CAA971103.1; -
CC SCD: S0003332; MDR1.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR004182; GRAM.
CC InterPro: IPR000195; RabGAP_TBC.
CC Pfam: PF00036; ehand; 1.
CC Pfam: PF02893; GRAM; 1.
CC Pfam: PF00566; TBC; 1.
CC SMART: SM00164; TBC; 1.
CC Transmembrane.
CC TRANSMEM 121 141 POTENTIAL.
CC TRANSMEM 398 418 POTENTIAL.
CC TRANSMEM 424 444 POTENTIAL.
CC SEQUENCE 950 AA; 109259 MW; 888BC930C1358657 CRC64;
KW Transmembrane.
FT TRANSMEM 121 141 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
SQ SEQUENCE 950 AA; 109259 MW; 888BC930C1358657 CRC64;

Query Match 4.5%; Score 274; DB 1; Length 950;
Best Local Similarity 23.3%; Pred. No. 4e-07;
Matches 98; Conservative 73; Mismatches 162; Indels 88; Gaps 12;

QY 714 KRTSRELRLWOKA-----ILQIILLRMEKE-----NOK-----743
DB 110 KRVERSPSEYEFALLVTLTYTGAKVLQIGIRYSEQFCDKLKNLKNIPNAKTLPAF 169
QY 744 LQASENDLNKRLKDYEEIT-----PCL-----KEVTYVWEKMLSTPGRSK 785
DB 170 LETSYSEFLIAKNILGKDDITPRAGLQGHFKYPGNPTVMVKEKAKRLWFDYFRENGRL 229
QY 786 IKFDMKMSAVCOGYPVRRHGRGEIWFLEAQHFKHQQFSPKQDPVYKELLQ---LT 842
DB 230 AVVQTPMFKRLIRIGVPMRMRGWEIWCSCGMYMANSGE-----YERILNENAGKT 282
QY 843 SQO-HAILIDLGRFTTHYFSAQAGOLSLYNILKAYSLLDQEVGYCQGLSFVAGILL 901
DB 283 SQAIDEERDLKSLPEYSAYQTEEGIORL--NVLTAISWKNPDVGYCQAMNIVVAGFL 340
QY 902 LHMSEEAFAKMLKFLMDFMGLRQYRPMIILQIQYQLSRLLDVYHRLYNHLEHETG 961
DB 341 IFMSEQAFCWCLNLC-DIVVPGYYSKTYGTLDDQRFESFVEDRMPVLWEYILQHDQ 399
QY 962 PSLYAAPWELTASQFPLGFVARVDMIFLOGTEVIFKVALSLGSHKPLIILQHENLET 1021
DB 400 LSVSLPWFSLFSTSMPLFAYRIMDIFFMNGSITLFOVALAVLKINADDLQADDGCM 459
QY 1022 IVDIFKSTLPNLG-----LVQMEKTIINQVFEMDIKQLOQAYEVEYHVLQEE 1067
DB 460 FIAIKHYFTLQSAHPDSSDKYKQITKQELLVTAEK-----EFSVISEE 507
QY 1068 L 1068
DB 508 M 508

RESULT 5

YN33_YEAST STANDARD; PRT; 633 AA.
AC P48566;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 73.0 kDa protein in CLAF-PUS4 intergenic region.
GN YNL293W OR N0470.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-S288C / FY1679;
RX MEDLINE-96132033; PubMed-8553702;
RA Maurer K.C.T., Urbanus J.H.M., Planta R.J.;
RT "Sequence analysis of a 30 kb DNA segment from yeast chromosome XIV
RT carrying a ribosomal protein gene cluster, the genes encoding a
RT plasma membrane protein and a subunit of replication factor C, and a
RT novel putative serine/threonine protein kinase gene.";
RL Yeast 11:1303-1310(1995).
CC 1- SIMILARITY: TO YEAST MIC1 AND S.POMBE SPAC4G8.04.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U23084; AAC49106.1; -
CC EMBL: U23084; AAC49106.1; -
CC SCD: S0005237; YNL293W.
CC InterPro: IPR000195; RabGAP_TBC.
CC Pfam: PF00566; TBC; 1.
CC SMART: SM00164; TBC; 1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 329 349 POTENTIAL.
CC TRANSMEM 422 442 POTENTIAL.
CC SEQUENCE 633 AA; 72999 MW; 5F62607B95BBAFB9 CRC64;

Query Match 4.1%; Score 247.5; DB 1; Length 633;
Best Local Similarity 23.2%; Pred. No. 5.7e-06;
Matches 78; Conservative 72; Mismatches 91; Indels 95; Gaps 15;

QY 738 ERENOKLQASENDLNKRLKDYEEITPCLKEVTYVWEKMLSTPG-----RSKIKFDM 790
DB 161 KQNNYISEAYD---KWWVEYSQY--CVRR-KHKWQLLEKSLPVYDDSPSRFSPKS 213
QY 791 ERMHSAVGSGVPRHHRGEIWFLEAQHFKHQQFSPKQDPV 832
DB 214 ERLKRYVRKGIPEWNGNAWHFARGQELKNKNGVYSLLRKMKQIKQNFQVQLD 273
QY 833 PYKELLKQLTSQOHAIDLGRFTTHYFSAQAGQ-----LSLYNLKAYSLLDQEV 887
DB 274 ---DIIEE-----DLNRTFPDNIHFQSSLNKKEGPIIKSLRRVLVAFSLNPKI 320
QY 888 GYCQGLSFVAGILLHMESEEAFAKMLKFLM-----FDMGLR-----KQYR 927
DB 321 GYCQSMNFTAGLLFLDEERAFWMLVIITSRYLPVGHINLEGVNIQDGLVLMCVKEYI 380
QY 928 PDMIIILQIQYQLSRLLDHYHRD-----LYNHLEEE-----IGP-SLYAAPWF 970
DB 381 P-----EWSYIIPSDIHQKNKNTFSPSNKVLFN-MQKNFYLRLPITLTASWF 432
QY 971 LTMFASQPLGFVARVDMIFLOGTEVIFKVALSL 1006
DB 433 MSCFVGVPVPIETTLRIWDCLFYEESHFLPKVSLAVL 468

RESULT 6

CENF_HUMAN STANDARD; PRT; 3210 AA.
ID CENF_HUMAN
AC P49454; Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENF-F kinetochore protein (Centromere protein F) (Mitosis) (AH
DE antigen).
GN CENPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

Qy 243 SLAFKELQDGLRSGFFSFEES-----DIENHLISGH-----NIVQPTDI 285
Db 962 EEAARQKLEKVTAEAKIKLEDDILVMDQNSKLSKERLLBERVSDLTNTLAAEEBK 1021
Qy 286 EENRMLFT-----IGQSEVILSPDTKKIALENKFRKISFCSGQIRHVDHFGICRESS 340
Db 1022 AKNTTKLSKSHESMISELEVLKKEKSQRQLEKLKRKL-----EG 1062
Qy 341 GGGGFHFVYVQCTNEALVDEIMMTLKQFTVAAVOQTAKAPQLCBGCPQLSLHKICE 400
Db 1063 DASDFH-----EQIAD-----LQAQIAELKQAKKEBEL-----QAALARLDE 1101
Qy 401 RIEGNSKTKL-ELQKHLTL-----TNOEQATIFEVQKLRPRNE-----441
Db 1102 EIAQNNALKKIRELEGHISDQLQDLDSERAARKAOKRDLGCELEALKTELEDTLDS 1161
Qy 442 -----QRENELISFLCLYEERKEHIHIGEMKQTSQMAAENIGSELPP-SATRP 491
Db 1162 TATQOELRAKREQVTV-LKKALDEETRSHEAQVQEMKOKHTQAVEELTEQLEQFKRAK 1220
Qy 492 RLDMKL-----NKARSL-----TESLESILSRGNKARGL-----521
Db 1221 NLDSKQTFLEKENADLAGELVLAQAKQEVHEKXKKLEVLQDLQSKCSDGERARAEUSD 1280
Qy 522 QEHSISVDLDS-----SLSTLSNTSKEPSVCKEALPISSESKF 561
Db 1281 KVHKLQNEVESVTGMLREAEAGKATLAKDVASLGSQLODQTLQOEETROKLNVTSK---1337
Qy 562 LIGSSEDLSSSESHLPREPAPLPQQAFFRANTLSHFPIECQPPQPARGSPGVORSK 621
Db 1338 -LRQLEDRNSLOQDLDEE---MEAKONLHRVSTLN---IQLSD-----SKKK 1379
Qy 622 LMRYHSVSTETPHE-----RKDFESKANHLGDSGTPVKTRRHSWR-----662
Db 1380 LQDFAS-TIEVMEBKKRLQKEMEGLSQOYEEKAAAYDKLEKTKNRLOQELDLVVDLDN 1438
Qy 663 -QQFLRVATQKACDS-----SSRYEDYSELGELPPRSPLEPCVEDGDFGPPPPPEK 713
Db 1439 QRLVSNLEKQKQKFDQALLAEKNISSYAD-----ER 1471
Qy 714 KRTSRELQKAI-----LQOJILLR--MEKENOKLOASENDLLNKR--LKLDYEBIT 764
Db 1472 DRAEAREKTKALSLARALEALEAKEELERTNKLKAMEDLVSSKDDVGRNVHLE 1531
Qy 765 PCLKEVTVWEKMLSTPGRSKIKFDMKMHSAVGQVPRHRGETWFKFLAQFHLKHQFP 824
Db 1532 KSKRALETQMEEM-----KTQLESEDDVQATEDAKLRLEV-----NMQALKGQFE 1577
Qy 825 SKQPKQVPYKELKQLTSQOHAILEDGRTFPHYPSAQLGAGQSLYLNILKAYSLLD 884
Db 1578 RDLQARDQNEEKRRQLOROLHEYETEL-----EDERQORALAAAKKLEGDLKDEL--1631
Qy 885 QEVGVCGLSFVAGILLHMSSEAFKMLFMDGMLRKQYRPMITLQIQYLSRL 944
Db 1632 -----QADSIAIK-----REAIKQ-----LRK-----LQAMKDFOREL 1661
Qy 945 HDYH--RD--LYNHLHEIGPSLYAPFWLTFMFPASQPLGFARVDFMIFLOQTEYIFK 1000
Db 1662 DDARASDEIFATSKENKAKSLEA-----DLMLQLEDLAAERAKQADKEELAE 1716
Qy 1001 VALSLGSHKPLIIOHE--NLETIVDFTKLPNLGLVQMCKTINQVEMDIAKOLQAYE 1058
Db 1717 LASSLSGRN--TLQDEKRRLEA-----RIAQLEEELE---EEQGNMEANSR 1758
Qy 1059 VEYHVLOEELSDSPSLDNQMDKLEKTNSSLRKNQDLLEQLQVANGRIQ-----S 1110
Db 1759 VRKATLQAEQUSNELATERSTAQNESARQQLERKNELSKLQEVGAEKAKLKSTVAA 1818
Qy 1111 LEATL-----EKLSSSKLQAMLTLELRSALIQTVLEELR-----1147
Db 1819 LEAKIALEQEQEAREKQKATSKLQKDKKLKLEVLQVDEDERKMAQYKQAEKGNTK 1878

Qy 1148 -----RRSAEPSDREPEC 1160
Db 1879 VKQLKQLEAEESQ 1895
RESULT 8
MYHB_HUMAN
ID MYHB_HUMAN STANDARD: PRT: 1972 AA.
AC P5749: 000396; P78422; O94944;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMHC).
GN MYH11 OR KIAA0866.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
human chromosome 16p and 16q";
RL Genomics 60:295-308(1999).
RN [2]
RP SEQUENCE OF 1-1266 FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirotsawa M.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro";
RL DNA Res. 5:355-364(1998).
RN [3]
RP SEQUENCE OF 885-1972 FROM N.A.
RX MEDLINE=93263189; PubMed=7684189;
RA Matsuoka R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,
Yanagisawa M., Masaki T., Takao A.;
RT "Human smooth muscle myosin heavy chain gene mapped to chromosomal
region 16q12.";
RL Am. J. Med. Genet. 46:61-67(1993).
RN [4]
RP SEQUENCE OF 1093-1972 FROM N.A.
RX TISSUE=Hippocampus;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL
ARTERY, BLADDER, ESOPHAGUS AND TRACHEA.
CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES.
CC -1- PFM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- INVERSION: A CHROMOSOMAL REARRANGEMENT, KNOWN AS PERICENTRIC
INVERSION INV(16)(P13022), PRODUCES A FUSION PROTEIN THAT CONSISTS
OF THE 165 N-TERMINAL RESIDUES OF CBF-BETA (PEPB2) WITH THE TAIL
REGION OF MYH11. THIS REARRANGEMENT IS ASSOCIATED WITH ACUTE
MYELOID LEUKEMIA OF M4EO SUBTYPE.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2). CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF001548; AAC31665.1; -
 DR EMBL; U91323; AAC35212.1; -
 DR EMBL; AB020673; BAA74889.1; -
 DR EMBL; D10667; -; NOT ANNOTATED_CDS.
 DR EMBL; X69292; CAA49154.1;
 DR HSSP; P08799; 1MMN.
 DR MIM; 160745; -
 DR InterPro; IPR000048; IQ.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 2.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00956; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family; Proto-oncogene; Chromosomal translocation.
 FT DOMAIN 1 785
 FT MYOSIN HEAD-LIKE.
 FT IQ.
 FT COILED COIL (POTENTIAL).
 FT CARBOXYL-TERMINAL.
 FT NP_BIND 178 185
 FT ACTIN-BINDING (BY SIMILARITY).
 FT ACTIN-BINDING (BY SIMILARITY).
 FT METHYLATION (TRI-) (POTENTIAL).
 FT ALKYLATION (SH-1) (POTENTIAL).
 FT ALKYLATION (SH-2) (POTENTIAL).
 FT ALKYLATION (SH-3) (POTENTIAL).
 FT EELK -> NSE (IN REF. 3).
 FT T -> S (IN REF. 3).
 FT KO -> NE (IN REF. 3).
 FT A -> S (IN REF. 4).
 FT T -> L (IN REF. 3).
 FT CONFLICT 1786 1786
 FT CONFLICT 1958 1958
 FT SEQUENCE 1972 AA; 227338 MW; 67665BB2AECE1277 CRC64;
 Query Match 3.6%; Score 217; DB 1; Length 1972;
 Best Local Similarity 19.8%; Pred No. 0.001;
 Matches 243; Conservative 169; Mismatches 441; Indels 374; Gaps 52;
 QY 119 KEDAV-----HROSICVYFADQTKVPEIISIRQAGKIARQEEELHCPSEFDDTSKKPE 174
 DB 857 KEDELQKTRQO-----KAENLEKEQKHSQLETKNLLQEQQAETEL---YAEAE 908
 QY 175 VLFPCGRTVAHKKAPALLDECIEFNHVSGRGSESPRPNPHAPTSQEPVPRPMK 234
 DB 909 M-----RVRLAAKK-----QLEELIHEWEARLEEE-----EDRQOQ---LQAEK 946
 QY 235 SFSQPL-----RSLAFKELQDGLRSSGFFSFEES-----DIENHLISGH----- 277
 DB 947 KMAQMLDLEEQLEEEAARQKLEKVAEAKIKKLEDEILVMDQNNKLSKERKLL 1006
 QY 278 -----NIVQPTDIENRML-----FTIGSEVYLISPDTKKIALEKNFKFPCSGG 325
 DB 1007 RISDLTTLNLAEEBEAKNLTKLNKSHESMISELEVLKKEKRSQLEKLRKL----- 1060
 QY 326 IRHVDHFGICRESSGGGFFHVCYVFOCTNEALVDEIMWTLKQFTVAVQOTAKAPAQ 385

Db 1061 -----EGDASDFH-----EQIAD-----LQAIQIAELKMLAKKEE 1091
 QY 386 LCEGCPLOSLHLKCRBIEGMSSTKYL-ELQKHLITL-----TQEQATIEE 431
 Db 1092 L-----QAALRLDDEIAQKNNAKIRELGHISDLQEDLSERARNKAERKQKRDGE 1146
 QY 432 EYQKLRPRNE-----ORENELIISFRLCYEERKQKHEHIGEMKQTSQMAA 477
 Db 1147 ELEALKTELEDLDTATQOELRAKREQEYTV-LKKALDEETRSHEAQVOEMRQHAQAV 1205
 QY 478 ENIGSELPSPATFRFLDMLKNK-----AKRSL-----TESL 508
 Db 1206 EELTEQL-EQFKRAKANLDKKNQTLKENADLAGELRVLGAKQVEHKKKKLEAQOVEL 1264
 QY 509 ESILSRGNKARGL---OEHSISVDLDS-----SLSSLTSLNTSKPE 545
 Db 1265 QSKSDGERARAEALNDKVKHKLQNEVESVTGMLNEAEGKAIKAKDVASLSSLODQTE-- 1322
 QY 546 SVCEKEALPISSESPKLLGSSDLSHPEPPAPLSPQOAFRRRANTISHPFIEQO 605
 Db 1323 -LQEBETROKLVSTKLRQLEERNS-LQDQLEDEE---MEAKONLERHISTLN---IQLS 1374
 QY 606 EPPQARGSPGVSKLMRYHSYSTETPERKDFESKANHLGDSGGTPV-----KTRR 658
 Db 1375 D-----SKKKLQDFASTVEALEBGGKRFQKEIENLTQOYEEKAAAYDKLEKTN 1423
 QY 659 HSWRO-----QIFLRVATPQKACDS-----SSRYEDYSELGELPPRSPLE 698
 Db 1424 RIQELDLVDLDNORQLVSNLEKKQKQKFDQALLAEKNISSYAD----- 1469
 QY 699 PVCEDGPPGPPPEKKRTSRRLRELWQKAI-----LQOILLR--MEKNQKLOASENDL 751
 Db 1470 -----ERDRAEAREKETRALSLARALEEALEAKEELERTNKLKAEEMDL 1516
 QY 752 LNKR--LKLDYEEITPCLKEVTVMEKMLSTGCRSKIKFDMKMHSAVQGVPRHHRGI 809
 Db 1517 VSKDDVGKNVHELEKSKRALETQMEEM-----KTQLELEDELOATEDAKLRLV 1567
 QY 810 WKFLAEQHLKHQFSPKQKQKDPVYKELKQTSQOHAIDLGRTPFTPHYFSAOLGAG 869
 Db 1568 -----NMQALKQGFERDLQARDEQNEERKRLQRLQHEVETEL-----EDERKORALAAA 1618
 QY 870 QLSLNLKAYSLLDOEVGYCGGLSFVAGILLHSEEAFAKMLAFMLFMDGLRKQYRPD 929
 Db 1619 KKKLEGLKLEL-----QADSAIKG-----REAAIKO-----LRK----- 1649
 QY 930 MIILQIQVLSRLSLHDYHR---DLVNLHEHE-IGPSLYAAPWFLTMAFSQPLGFVAR 985
 Db 1650 ---LQAKMKDFORELADARSDEIFATAKENKAKSLEA-----DLMLQEDLAAAE 1701
 QY 986 VFDMIFLQGTVEVIFKVALSLGLSHKPLILQHE--NLETTIVDFIKSTPLNGLVQMEKTN 1043
 Db 1702 ARKQADLEKEELAELASSLSGRN---ALQDEKRRLEA-----RIAQLBEELE 1746
 QY 1044 QVFEMDIAKQAYEYEHVLEQELIDSSPLSDNQMDKLEKTNSSLRKNLDLLEQLQV 1103
 Db 1747 ---EEQGNMEASDRVRKATQQAQOELSNELATERSTAQKNESARQOQLERKSLKHE 1803
 QY 1104 ANGRQOS-----LEATI-----EKLSSSESKLKQAMLTLELERS 1137
 Db 1804 MEGAVKSKFKSTIALEAKIAQLEQVEQEAQKQATKSLQKOKKLEILLQVEDERK 1863
 QY 1138 ALLQVTEELRRRSAPFSDEPECTOPE 1164
 Db 1864 MAEQYKEQAEKNARVKQLKQLEBAE 1890
 RESULT 9
 MYHB_CHKC
 ID MYHB_CHKC STANDARD; PRT; 1978 AA.
 AC P10587;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)
 Myosin heavy chain, gizzard smooth muscle.
 Gallus gallus (Chicken).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 Gallus.
 NCBI_TaxID=9031;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=88118918; PubMed=2892941;
 Yanagisawa M., Hamada Y., Katsuragawa Y., Imamura M., Mikawa T.,
 Masaki T.;
 "Complete primary structure of vertebrate smooth muscle myosin heavy
 chain deduced from its complementary DNA sequence. Implications on
 topography and function of myosin.";
 J. Mol. Biol. 198;143-157(1987).
 [2]
 REVISIONS.
 Masaki T.;
 Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE OF 1-203.
 MEDLINE=88032919; PubMed=3312184;
 Maita T., Onishi H., Yajima E., Matsuda G.;
 "Amino acid sequence of the amino-terminal 24 kDa fragment of the
 heavy chain of chicken gizzard myosin.";
 J. Biochem. 102;133-145(1987).
 [4]
 X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-818, AND REVISIONS 204-215.
 MEDLINE=98412652; PubMed=9741621;
 Dominguez R., Freydon Y., Trybus K.M., Cohen C.;
 "Crystal structure of a vertebrate smooth muscle myosin motor domain
 and its complex with the essential light chain: visualization of the
 pre-power stroke state.";
 Cell 94;559-571(1998).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 EMBL: X06546; CAA29793.1;
 PIR: S03166; S03166.
 PDB: 1BR1; 09-SEP-98.
 PDB: 1BR2; 09-SEP-98.
 PDB: 1BR4; 09-SEP-98.
 InterPro: IPR000048; IQ.
 InterPro: IPR004009; Myosin_N.
 InterPro: IPR002928; Myosin_tail.
 InterPro: IPR002017; Spectrin.
 InterPro: IPR001609; myosin_head.
 Pfam: PF00612; IQ; 1.
 Pfam: PF02736; Myosin_N; 1.

DR Pfam: PF01576; Myosin_tail; 1.
 DR PRINTS: PRO0193; MYOSINHEAVY.
 DR PRODOM: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSc; 1.
 DR PROSITE: PS00096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Calmodulin-binding;
 KW Multigene family; 3D-structure.
 FT INIT_MET 0
 FT DOMAIN 1 790 MYOSIN HEAD-LIKE.
 FT DOMAIN 791 820 IQ.
 FT DOMAIN 849 1978 RODLIKE TAIL (S2 AND LHM DOMAINS).
 FT DOMAIN 849 1978 COILED COIL (POTENTIAL).
 FT NP_BIND 176 183 ATP.
 FT DOMAIN 666 688 ACTIN-BINDING.
 FT DOMAIN 767 781 ACTIN-BINDING.
 FT MOD_RES 1 1 BLOCKED.
 FT MOD_RES 127 127 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 706 706 ALKYLATION (SH-1).
 FT MOD_RES 716 716 ALKYLATION (SH-2).
 FT CONFLICT 127 127 MISSING (IN REF. 3).
 FT CONFLICT 204 215 KDTISITQPSFS -> RTPASLKVHLFP (IN REF. 1).
 SQ SEQUENCE 1978 AA; 228663 MW; B7B6C923E5273D93 CRC64;
 Query Match 3.5%; Score 212.5; DB 1; Length 1978;
 Best Local Similarity 19.2%; Pred. No. 0.0017;
 Matches 191; Conservative 147; Mismatches 310; Indels 345; Gaps 43;
 QY 356 NEALVDIEMTLKQA-----FTVAAGV-OTAKAPALQCEGCP 392
 DB 1037 HESMISELEVLKKEEKSQBLEKIKKLEGGESDLHQIAELQALQIAELKQALAK--KE 1094
 QY 393 QSLHKLCEIRIEGMNSKTKL-----ELQKHLTTL-----TNQQAIFEEVQ 434
 DB 1095 EELQAALARLEDEFSQNNALKKIRELESISDLQEDLESEKAARNKAQKRDLSLEE 1154
 QY 435 KLRPNE-----ORENELIISFLRCYEEKQKEHIHIGMKOTSMAANI 480
 DB 1155 ALKTELEDLTDTTATQOELRAKREQVTV-LKRALEEETRTHEAQVQEMRKHTQAVBEL 1213
 QY 481 GSELP--SATFRIDMLKNKAK-----RSITES-----LESIL 512
 DB 1214 TEQLEQFKRAKANDKTKQTLKONADIANEIRLSQAKQDVHEKKKLEVLQDLQSKY 1273
 QY 513 SRGNKAR-GLQF--HSISVDLSDSLSTLSNTSKEPSYCEKEALPI-----SE 557
 DB 1274 SDGERVTELNEKVHKLQIEVE-NVTSLLNEAESKNIKLTKDVATLGSQLODTQELLQEE 1332
 QY 558 SSFKL-----LGSSEDLSSSESHLPPEAPLSPQQAQFRFRANTLSHFPIECQPPQPAR 612
 DB 1333 TROKLNTVTKLRQLEDKNSLQELDEE---VEAKONLERHISTLT---IQLSD----- 1380
 QY 613 GSPGVSRKLMRYHSVSTETPERK-----DFESKANHLGDSGGTTPVKTRRH- 659
 DB 1381 -----SKKKLQEF-TATVETMEEEKKLQRIEISLTQOFEEKAASYDKLETKRKLQEL 1434
 QY 660 -----SWRQIFLRVATPQKACDS-----SSRYEDYSELGELPPRSPLEPCVCDG 704
 DB 1435 DDLVVDLDNQRLVSNLEKKQKFDQMLAEKNISSYAD----- 1474
 QY 705 PFGPPPEKKRTSRELRLWQKAI-----LQOILLR--MEKENQKLOASENDLLNKR-- 755
 DB 1475 -----ERDRAEAAREKETKALSARALEALEAKEELERTNKLKAEMEDLVSSKDD 1527
 QY 756 -----LKLDEEITPCLEKVTTWKMLSTPGRSKIKFDMKMKHSAVGGV 801
 DB 1528 VGKVVHELESKRTLEQQVEEMKTLQEELED--ELQAAEDAKLLEYNQAM----- 1577
 QY 802 PRHRGEIWKFLAEQFHLKHQFPSPKQPKDVPYKELKQLTSQOHALIDLGRFTPHY 861
 DB 1578 -----KSQFERDLQARDEQNEKRRLKQLKQHEHETEL-----EDERK 1615

QY 827 -OOPKDVVKELLKOLTQSQAAILDLGRFTFTHPVFSAQAGQLSLYNILKAYSLILDO 885
 Db 649 CQEKETLLKD--KEIIPAHEEWN-EKTEKLDVKQTELESLSSEVLKARKHLEE 705
 QY 886 FVGYCOGLSFVAGIILLHMSSEAFKMLKFLMDMLGRK--QYRPMII----- 932
 Db 706 E-----LSVLKQDTDKMKQLEAKMDEQKNHHQOQQVDSIIKEHVSIGRT 750
 QY 933 ---LOIQMYQLSRLLDHYRDLYNHLHEHEI-----GFSLYAAPFLTFWFS- 976
 Db 751 EKALDQINOELLKE--RD--KRLKEHQAHVENLEADIKRSEGELOQASAKLDVFOSY 806
 QY 977 QPPLGFVARVFMIFLQGTVEVLFKVALSLGSHKPLIL-----OHENLETIVDFI 1026
 Db 807 QSATHEQKAYE-----EQLAQLOQLDLLETRILLTKOVAEVAQKDVCTELDAH 859
 QY 1027 KSTPLNLGLVQMEKTIQVNFENDIAQIAQYAEVYHVLQEEILDSPLSDNQMDKLEKT 1086
 Db 860 KIOVQDL-MQOLEKQ-NSEMEQKVKSLTQVYE-----SKLEDGNKEQEQTKQILVEKE 910
 QY 1087 NSSL-----RKQNLDLLEQ-----LOVANGRIQSLBATI 1115
 Db 911 NMILQMRGQKKEIEILTKQSAKEDSIHLNEEYETKFKNOEKMKVKQKAKEMQETL 970
 QY 1116 -EKLSSSESKLQAMLTLELERSALLQTVVEELRRRSABPSDREPTCTOPEPTG 1167
 Db 971 KKLLDQEAFLK-----ELENTAL-----ELSQREKQFNKMLEMAQANSAG 1013

RESULT 11
 CENE_HUMAN
 ID CENE_HUMAN STANDARD; PRT; 2663 AA.
 AC 002224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENPE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis.";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor.";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL; Z15005; CAA78727.1; -;
 DR PIR; S28261; S28261.
 DR HSSP; P17119; 3KAR.
 DR MIM; I17143; -;
 DR InterPro: IPR001752; kinesin.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335
 FT DOMAIN 336 2471
 FT DOMAIN 2472 2663
 FT NP_BIND 86 93
 FT NP_BIND 93 93
 SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Query Match 3.4%; Score 207; DB 1; Length 2663;
 Best Local Similarity 17.8%; Pred. No. 0.005;
 Matches 160; Conservative 165; Mismatches 290; Indels 284; Gaps 34;
 QY 359 LVDEIMMTLKQAFVAAVQQTAKAPALCEGCPQLQ-----S 394
 Db 913 ITEKLOQTLEEVKTLTQEKDLDL---QLOESLIQIRDLKSDIHTVNNIDTQSLRNA 969
 QY 395 LHKLCERTEGMSKTKL--ELQKHLTLTN-----QEQTAFEEYVKLRPRNEQ-- 442
 Db 970 LESLQHQETINTLAKSISEEVRNLHMEENTGETKDFQOKMVGIDKKDLEAKNTQL 1029
 QY 443 ---RENELITSFLRCLYEKQKEHIGEMKQTSQMAAENSGELSPSATFRDLMLKN 498
 Db 1030 TADVKNDEII-----EQQRKIFSLQEKNLQOMLESVIAE-----KE 1067
 QY 499 KAKRSLSLSELSRGNKARGL-----QHSISVDLDSLSSTLSNTSKEPS 546
 Db 1068 QKTDLKENIEMTIENQELRLGLDELAKQOEIVAQERHAIKKEGELSRCDRLAEVEE 1127
 QY 547 VCEKEALPISSESKLGSSEDLSS-----DSHSLPEEPAPLSPQQAFFRRANTLSHP 601
 Db 1128 KLKESQQLQEQKQOLLNVQEMSEMOKKINEIENKNE-----LKNKELTLEHME 1178
 QY 602 IECQEPPOPARGSPGVSRKMLRYHSVSTPHER-----KDFESKANHL-----GDSGG 651
 Db 1179 TERLE-----LAQKNENYEEVKSITKERVKLKELQKSFETRDHLRGYIREIA 1228
 QY 652 TPVKTRRHSWRQQIFLRVATPQACDSSSRVEDYSELGELPPRSPLEPCEDGPGPPPE 711
 Db 1229 TGLQTEELKTAHILK-----1245
 QY 712 EKKTSRELRELWQKAILQIQLLLMRKENOKLQASENDLLNKRKLQDYETTPCLKEVT 771
 Db 1246 EHOETIDELRRSVSEKTAQIINTQDLKSHTKLQ-EIIPVLHEE-----QELLPNVKVS 1299
 QY 772 TWKEM-----LSTPGRSKIYFDMKMSHSAVQGVPRHHRGEIWKFLAOPFLKHQFSPK 826
 Db 1300 ETQETMNELELLTEQSTTKDSTTLARIE-----MERLNLKFKQES 1340
 QY 827 QQPKDVPYKELLKOLTQSQAAILDLGRFTFTHPVFSAQAGQLSLYNILKAYSLILDOE 886
 Db 1341 Q-----EIKSUTKERNLKTKEALEVHKDOLKEHI---RETLAKIQESQKQSQS 1389
 QY 887 VGYCOGLSFVAGIILLHMSSE--EAFKMLKFLMDFMDGLRKYRPPD---MILQIQMQLS 941

Db 1390 -----LNMEKDNETHIVSEM-----EQFKPKDSALLRIEIMLGSL 1427
 QY 942 RLLHDYRDLYNLEEHGIPSLYAAWFLTMFASOPFLGFVARVDFMIFLQGTIEYFKV 1001
 Db 1428 KRLQESHDEMKVAKED-----DLQRLQ--EVLQSE 1457
 QY 1002 ALSGLSHKPLILQHNLETFIDVFIKSTLPNLGLVOMEKTIQOV-----FEMDIA---K 1052
 Db 1458 SDQLKENIKRIVAKH--LETEEL--KVAHCCLKKEQETINELRVNLSEKETEISTQIK 1512
 QY 1053 OLQA-----YEYEHVQLBELDSSPLSDNQMDKLEKTNSSLRKQNLDLLE- 1099
 Db 1513 QLEAINDKLNKTOEYIEKEQENIKOI---SEVOENVELKQFKHRKAKDSALOSIES 1569
 QY 1100 -QLQVANGRIQSLEATIEKLLSSSEKSLKQAMLTLELERSALLQTVLEELRRRSAPSDRE 1157
 Db 1570 KMLELTLN-RLQESQEEIQIMKEEMKRVQEAQLQIEROOLKENTKEIYAKMKESQKE 1627
 RESULT 12
 MYHB_RABIT
 ID MYHB_RABIT STANDARD; PRT; 1972 AA.
 AC P35748;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, smooth muscle isoform (SMHHC).
 GN MYH11.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_Taxid:9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92073350; PubMed-1961735;
 RA Babji P., Kelly C., Periasamy M.;
 RT "Characterization of a mammalian smooth muscle myosin heavy-chain
 gene: complete nucleotide and protein coding sequence and analysis of
 the 5' end of the gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).
 CC 1- FUNCTION: MUSCLE CONTRACTION.
 CC 1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC 1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC 1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC 1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC 1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC 1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC 1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M77812; AAA31395.1;
 DR FIR; A41604; A41604.
 DR HSSP; P08799; LMMD.
 DR InterPro; IPR000048; IQ.
 DR InterPro; IPR004009; Myosin.N.
 DR InterPro; IPR002928; Myosin.Tail.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001609; myosin_head.

DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00663; myosin_head; 1.
 DR Pfam; PF02736; Myosin.N; 1.
 DR Pfam; PF01576; Myosin.Tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS50096; IQ; 1.
 KW Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family. MYOSIN HEAD-LIKE.
 FT DOMAIN 1 784 IQ.
 FT DOMAIN 785 807
 FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
 FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
 FT NP_BIND 178 185 ATP (POTENTIAL).
 FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 763 777 ACTIN-BINDING (BY SIMILARITY).
 FT MOD_RES 129 129 METHYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 701 701 ALKYLATION (SH-2) (POTENTIAL).
 FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1972 AA; 227318 MW; 2061A224288D6A4C CRC64;
 Query Match 3.3%; Score 202; DB 1; Length 1972;
 Best Local Similarity 19.8%; Pred. No. 0.0062;
 Matches 187; Conservative 127; Mismatches 352; Indels 278; Gaps. 37;
 QY 377 QQTAKAPALCE-----GCPLASLHKLCERIEGMNSKTKL-----ELQKHLTTL----- 421
 Db 1069 EQIADLQAQIAELKMLAKKEELQAALARLEDETSQKNALKKIELEGHISDLQEDLD 1128
 QY 422 -----TNOQATIFEVVKLRPRNE-----ORENELIISFLRCLYEKK 459
 Db 1129 SERAARKAEKQKRDGLGELEALKTELEDTLTATTQQLERAKREQEVTV-LKKALDEET 1187
 QY 460 QREHIHIGEMKQTSMAAENIGSELPP-SATFRFLDMLK-----NKAQ 501
 Db 1188 RSHEAQVQEMRKHTQVVEELTEQLQFKRAKLANLTKQTLEKENADLAGELRVLGQAK 1247
 QY 502 RSL-----TESLEISLGNKARGL---QEHISIVDLDS----- 532
 Db 1248 QVEVHKKKKLEVLQELQSKSDGERARAEALNDKVKHLQNEVESVTGMLSEABGKAUKA 1307
 QY 533 ----SLSSTLSNTSKPSVCEKEALPISESSFKLIGSSEDLSDSDSHLPEEPAPLSQ 588
 Db 1308 KEVASLGSQLDQTQELLQETKQLNVSTK----LQLEDESNLSQEQLEDE---MEAKQ 1360
 QY 589 AFRRRANTLSHFPIECQEPQPARGSPGVQSQRKLMRYHSV--STETPHER--KDFESKAN 644
 Db 1361 NLERHISTLN---IQLSD-----SKKKLQDFASTVESLEBGRKRFQKEIESLTQ 1406
 QY 645 HLGDSGGTPVKTRRHSWR-----QQIFLRVATPQKACDS-----SSR 681
 Db 1407 QYEEKAAAYDKLEKTKNRLQQLDLDLVLDNQQLVSNLEKKQKFKFDQLLAEKNISSK 1466
 QY 682 YEDYSELGELPPRSPLEPCVEDGPPGPPPEKKRTSRERELWQKAI-----LQQLILLR 736
 Db 1467 YAD-----ERDRAEAAREKETKALSARALEALEAK 1499
 QY 737 --MEKNQKLQASENDLLNKR--LKLDYBEITPCLKEVTVVWEKMLUSTPGRSRIKFDMEK 792
 Db 1500 EELERTNKLKAAEMEDLVSSKDDVGKGVHLEKSKRALETQMEEM-----KTQLEE 1550
 QY 793 MHSVAGQGVPRHRRGIWFLAEQFHLKHQFSPKQPKDVPVYKELLKQLTSQOHAIDL 852
 Db 1551 LEDELQATEDAKLRLEV-----NMQALKVQFERDLQARDEQNEEKKRRQRLQRLHETEL 1605
 QY 853 GRTFPTHYFSAQLGASQLSYNLKAYSLLDQEVGCGQLSFVAGILLHMHSEEFAMK 912
 Db 1606 ----EDERQKRALAAAKKKLGSLDKDEL-----QADSAIKG-----REEAIKQ 1646

```
QY 913 LKFLMFDMLGRKQYRPMIIIOIQMYQLSRLLHBYH---DLYNHLBEHE-IGPSLYAAP 968
Db 1647 L-----LKLQAKMDFQRELEDAARSDFIFATAKENERKAKSLEA-- 1687
QY 969 WFLTMFASQFPLGFVARVEDMIFLQGVTEFKFVALSLGSHKPLILOHE--NLETIVDFI 1026
Db 1688 ---DLMQLQEDLAAAEARKQADLEKELEBELASSISGRN---ALQDEKRRLEA----- 1736
QY 1027 KSTLPNGLVOMEKTIINOVFENDIAKQLQAYEVEYHVLQELIDSSPLSDNORMDKLEKT 1086
Db 1737 -----RIAQLEBELE---EEQCNMEAMSDRVKATQQAQBSLQNELATERSTAKNESA 1786
QY 1087 NSSLRKNLDLLEQLQVANGHIOQS-----LEATI-----EKLKS 1120
Db 1787 RQQLERONKELSKLQMEGAVKSKFKSTIAALEAKIAQLEEQVEQAREKQAQAALKQ 1846
QY 1121 SESKLQAMLTLELERSALQTVBELARRRSAPESDRPECTQPE 1164
Db 1847 RDKLKLKMLQVEDERKMAEQYKEQAEGKNAKVQLKRLQLEAE 1890
RESULT 13
MYHA_BOVIN STANDARD; PRT; 1976 AA.
AC Q2791;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
DE type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
GN MYH10.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RA SEQUENCE FROM N.A.
RA Ohara M., Ishiguro N., Shinagawa M.;
RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SEQUENCE OF 204-302 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=95301542; PubMed=7782316;
RA Itoh K., Adelstein R.S.;
RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle
RL myosin heavy chain II-B.";
RL J. Biol. Chem. 270:14533-14540(1995).
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
CC CAPPING (BY SIMILARITY).
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AB022023; BAA36494.1;
CC EMBL; U15716; AAA87715.1;
CC HSP: P08799; IIVK
DR InterPro; IPR000048; IQ.
```

```
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
FT DOMAIN 786 815 IQ.
FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP (POTENTIAL).
FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1976 AA; 229097 MW; 6144354451C0F790 CRC64;

Query Match 3.38; Score 197; DB 1; Length 1976;
Best Local Similarity 19.58; Pred. No. 0.011;
Matches 193; Conservative 127; Mismatches 318; Indels 354; Gaps 39;

QY 357 EALVDEIMTMLKQAFVAAVQOTAKAPALCEGCPL-----QSLHK----- 397
Db 1075 QAQIDELKI-----QVAKKEEL-QGALRGDDDELHNNALKVVRLOAQIA 1121
QY 398 -LCERIEGMNSKTKYLELQK-----HITLTNQEQATIFEV--OKLPRNQRENELII 449
Db 1122 ELQEDFSEKASRNKAEKQKRDLSLEALKTELEDITLDTAAQQLTKRKEQVAEL-- 1179
QY 450 SFLCLYEEKQKEHIHIGEMKQTSQMAAENIGSELPSATFRDLMLKKA----- 500
Db 1180 --KKALEETKSHBAQIQDMQRHATALEELSEOL-EOAKRFKANLEKNKOGLETDNKL 1236
QY 501 -----KR-----SLTESLESILSRGNKARGLOEHSISVDL---DSSLST 537
Db 1237 ACEVKVLQVKAESEHKRKLDAQVQELHAKVSEGDRLR-----VELAEKANKLQNE 1288
QY 538 LSNWTSKEPSVCEKEALPISSEFKLLGSSDLSSDSHLPPEEPAPLSPQOAFRRANTL 597
Db 1289 LDNVSTLLEAEKGIKFAKDAAGL---ESQLQDTQELQEEET-----ROKLNLS 1335
QY 598 SHRP-----TECOBPQOPARGSPGVORSOKLMRYHVSSTETPHE----- 635
Db 1336 SRIRQLSEERSLSQEQEEEEEARRS---LEKQLAQALQALDTYKKVDDDLGCTIENLEE 1392
QY 636 -----RKDFESKANHLGDSGT-----PVKTRRH-----SWRQOIFLRVATPQKA 675
Db 1393 AKKLLKLDVEVLSQLSEKALAYDKLEKTKRLQQLQELDLDDLVDLHQRIVSNLEKKQK 1452
QY 676 CDS-----SSRYEYSELGELPPRSPLEPCVEDGFGPPPEKKRTSRELRLWKA 727
Db 1453 FDQLLABEKNISARY-----AEERDRAEAAREKETKA 1485
QY 728 I-----LQOIILLR--MEKENOKLQASENDLLNKR-----LKLDYEET 764
Db 1486 LSLARALEALEAREEAERQNKQURADMEDLMSSKDDVGKVNHELEKSKRALEQVEEMR 1545
QY 765 PCLKEVTVMKMLSTPGRSKIKFDMKMHSAVGQVPRHHRGEIWKFLABQFHLKHQFP 824
Db 1546 TQLEED--ELQATEDAKURLEVMQAM-----KAQFE 1577
QY 825 SKQPKDVPYKELKQITSQOHAILEDGRTFPHYFSAQLGAGQLSYLILKAYSLLD 884
Db 1578 RDLQTRDEQNEEKRLIKQVRELEAELEDEKORALAVASKKMEIDLKDL----- 1629
QY 885 QEVCYCGGLSFVAGILLHMHSEEAFFKMLKFLMDFMDGLRKQYRPMIIIOIQMYQLSRLL 944
```

Db	1630	-----EADIERANKARDEVIQ-----LQRM-----	1654
Qy	945	HDYHRDLYNHLHEEIGPSLYAAPWFLTMFASQPLGFAVRFDMIFLQGTPE-----	996
Db	1655	KDYQRE---LEEAR-----ASR-----DEIPAQSESEKKLKS 1685	
Qy	997	--VTFKVALSLLGSHKPLIIOHENLETIVDFIKTLPNLGLVQMEKTNQVFMEDIAKOL 1054	
Db	1686	EAEITLQOEELASSERRARRHAEQERDELADIAANSAGSKSALLDEK---RRLEARIAQLE 1742	
Qy	1055	QAYEVEYH-----VLOEELDSSPLSDNQMDKLEKTNSSLRKONLDLLEQL 1102	
Db	1743	EELEEGOSNMELNDRFRKTTLOVDTLTNTELAASRAAQSDNARQOLERONKELKAKLO 1802	
Qy	1103	VANGRIOS-----LEATEKL-----LSESKLQAMLTLELR 1136	
Db	1803	ELEGAVSKFKATISALEAKIGLEEQEQAKERAANKLVRTTEKLEIFMQVEDER 1862	
Qy	1137	SALLQTVELRRRSAPSD--REPECTOPEPT 1166	
Db	1863	RHADQYREQMEKANARKQLKROLEEEAAT 1894	
RESULT 14			
TMF1_HUMAN			
ID	TMF1_HUMAN	STANDARD;	PRT; 1093 AA.
AC	P82094;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	TATA element modulatory factor (TMF).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND FUNCTION.		
RC	TISSUE=Cervical carcinoma;		
RX	MEDLINE=93028466; PubMed=1409643;		
RA	Garcia J.A., Ou S.-H.I., Wu F., Lusis A.J., Sparkes R.S., Gaynor R.B.;		
RT	"Cloning and chromosomal mapping of a human immunodeficiency virus 1		
RT	"TATA" element modulatory factor.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 89:9372-9376(1992).		
CC	-1- FUNCTION: THIS PROTEIN BINDS THE HIV-1 TATA ELEMENT AND INHIBITS		
CC	TRANSCRIPTIONAL ACTIVATION BY THE TATA-BINDING PROTEIN (TBP).		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed, usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; L01042; RAD54608.1;		
DR	MM; 601126;		
KW	Transcription regulation; DNA-binding; Repressor; Coiled coil.		
FT	DOMAIN 439 922	COILED COIL (POTENTIAL).	
FT	DOMAIN 984 1092	COILED COIL (POTENTIAL).	
SQ	SEQUENCE 1093 AA; 123170 MW; 261338ESF4677BE CRC64;		
Query Match 3.28; Score 196.5; DB 1; Length 1093;			
Best Local Similarity 17.99; Pred. No. 0.0055;			
Matches 196; Conservative 181; Mismatches 362; Indels 353; Gaps 46;			
Qy	205	GSRGSESRPNPPHAAPTGSGEYVRRMRKFSQPLRSLAFRKLQDGLRSGGFTSSF 264	
Db	63	GLKNTPEQ-SPTASPKATIKPVRTVVD-----ESENFFSAF 100	
Qy	265	EESDIENHLISGHNIQVPTDIEENRMTLTIGQSEVYLISPTDKIALERKFKISFCSQ 324	

RESULT 15

MYH9_HUMAN STANDARD; PRT; 1960 AA.
 AC P35579; G60805;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE MYOSIN heavy chain, nonmuscle type A (Cellular myosin heavy chain,
 DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
 GN MYH9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
 RA Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
 RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
 RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
 RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
 RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
 RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
 RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
 RA Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
 RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kersey J.,
 RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
 RA Lloyd C., Lloyd D.M., Marlyn I.D., Mashreghi-Mohammadi M.,
 RA Matthews L., Mccann O.T., Mcclay J., McLaren S., McMurray A.A.,
 RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
 RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
 RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
 RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
 RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
 RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
 RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
 RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
 RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
 RA Mitsuyma S., Roe B.A., Chang F., Chu L., Crabtree J., Deschamps S.,
 RA Do A., Do T., Dorman A., Fan F., Fu Y., Hu P., Hua A., Kenton S.,
 RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.P., Loh P., Mala E.,
 RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
 RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
 RA Wu H., Yao Z., Zhan M., Zhang G., Chissole S., Murray J., Miller N.,
 RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
 RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
 RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
 RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K.,
 RA Nelson J., Korf I., Bedell U.A., Hillier L., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tiliakou Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [2]
 RP SEQUENCE OF 1-1337 FROM N.A.
 RX MEDLINE=92003925; PubMed=1912569;
 RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,
 RA Arnaout M.A., Clayton L.K., Tenen D.G.;
 RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'
 RT cDNA clones, characterization of the protein, chromosomal
 RT localization, and upregulation during myeloid differentiation.";
 RL Blood 78:1826-1833(1991).
 RN [3]
 RP SEQUENCE OF 1-715 FROM N.A.
 RX MEDLINE=91316803; PubMed=18660190;
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
 RA Guila D., Adelstein R.S., Weir L.;
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located

on different chromosomes";
 Circ. Res. 69:530-539(1991).
 RN [4]
 RP SEQUENCE OF 714-1960 FROM N.A.
 RX MEDLINE=90138958; PubMed=1967836;
 RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;
 RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
 RT through alternative polyadenylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
 RN [5]
 RP VARIANT DFNA17 HIS-705.
 RX MEDLINE=20489856; PubMed=11023810;
 RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Matre A.N.;
 RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
 RT nonmuscle myosin MYH9.";
 RL Am. J. Hum. Genet. 67:1121-1128(2000).
 RN [6]
 RP VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.
 RX MEDLINE=20428192; PubMed=10973259;
 RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
 RA Grigori G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
 RA Iolascon A., Zelante L.L., Savoia A., Balduini C.L., Noris P.,
 RA Magrini U., Belletti S., Heath K.E., Babcock M., Gluckman M.J.,
 RA Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:103-105(2000).
 RN [7]
 RP VARIANTS MHA ILE-1155 AND LYS-1841.
 RX MEDLINE=20428193; PubMed=10973260;
 RA Kelley M.J., Javien W., Ortel T.L., Krczak J.F.;
 RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
 RT May-Hegglin anomaly.";
 RL Nat. Genet. 26:106-108(2000).
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CC CAPPING.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
 CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME
 CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC WITH ADDITIONAL ALPOPT-LIKE CLINICAL FEATURES OF SENSORINEURAL
 CC DEAFNESS, CATARACTS AND NEPHRITIS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
 CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS
 CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
 CC COCHLEOSACULAR DEGENERATION.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 282215; CAB05105.1; -;
 DR EMBL; M81105; AAA5988.1; -;
 DR EMBL; M69180; AAA61765.1; -;

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:22:27 ; Search time 29.58 Seconds
(without alignments)
3794.197 Million cell updates/sec

Title: US-09-762-311-5

Perfect score: 6055

Sequence: 1 MEPIFTARKHLLPNEVSD.....RSAEPSDREPECTQPEPTGD 1168

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4976	82.2	1141	2 T29104	Tbcl protein - mou
2	2828.5	46.7	1299	2 T00261	hypothetical prote
3	1462.5	24.2	1379	2 T13718	pollux gene protei
4	580	9.6	1030	2 T13163	Rab6 GTPase activa
5	523.5	8.6	1142	2 G89453	protein F35H12.2 (
6	465	7.7	894	2 S61015	hypothetical prote
7	421	7.0	876	2 T49801	hypothetical prote
8	397	6.6	480	2 T29370	hypothetical prote
9	387	6.4	1031	2 T38411	probable GTPase ac
10	386	6.4	356	2 T51376	plant adhesion mol
11	343	5.7	772	2 S62481	hypothetical prote
12	317.5	5.2	720	2 S50920	hypothetical prote
13	317	5.2	814	2 T47841	hypothetical prote
14	293	4.8	519	2 T16712	hypothetical prote
15	290	4.8	468	2 T48686	hypothetical prote
16	282	4.7	834	2 T39891	probable integral
17	276	4.6	458	2 G88391	protein R06B10.5 (
18	275.5	4.5	808	2 T25748	hypothetical prote
19	274	4.5	950	2 S64405	MIC1 protein - yea
20	272.5	4.5	419	2 T08683	hypothetical prote
21	270	4.5	1244	2 T19165	hypothetical prote
22	258.5	4.3	1270	2 T26720	hypothetical prote
23	252.5	4.2	528	2 S51887	hypothetical prote
24	247.5	4.1	492	2 S51885	probable membrane
25	247.5	4.1	633	2 S60408	probable membrane
26	243.5	4.0	487	4 S51886	hypothetical prote
27	234	3.9	637	2 S66953	hypothetical prote
28	232	3.8	856	2 T34491	hypothetical prote
29	227	3.7	1938	2 JC5421	smooth muscle myos

RESULT 1

T29104

Tbcl protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T29104

R:Richardson, P.M.; Zon, L.I.

Oncogene 11, 1139-1148, 1995

A:Title: Molecular cloning of a cDNA with a novel domain present in the tre-2 oncogen

A:Reference number: Z20569; MUID:96032578

A:Accession: T29104

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1141 <RIC>

A:Cross-references: EMBL:U33005; NID:g988220; PID:g988221; PIDN:AAA85223.1

C:Genetics:

A:Gene: tbcl

Query Match

82.2%; Score 4976; DB 2; Length 1141;

Best Local Similarity 87.9%; Pred. No. 9.2e-270;

Matches 970; Conservative 47; Mismatches 74; Indels 12; Gaps 5;

QY 37 MPMLPWVVAEVRRLSRGSTRKEPTKQVRLCVSPSGRLCEPEPGRSQOQDPLIYSSIFEC 96

Db 1 MPMLPWVVAEVRRLSGGCSKKEPTKQVRLWVSPSGRLCEPEPDLKESQFWDPLICSSIFEC 60

QY 97 KPORVHKLIHNSHDPVSFACLIKEDAVHROSICVFKADDTQKVPEIISIRQAGKIARQ 156

Db 61 KPORVHKLIHNSHDPVSFACLIKEDAAHRQSLCVFKADDTQKVPEIISIRQAGKIARQ 120

QY 157 EELHCPSEFDDTFSKKFEVLFCGRVTVAAHKKAPPALIDECIEKFNHVSGRGSSSRPNP 216

Db 121 EELHCPSEFDDTFAKKEVLFCGRVTVAAHKKAPPALIDECIEKFNHVSGRGRTD----- 174

QY 217 PHAATGSG-DEPVRRPNRKSFSGPLRLSLAFRLQDGLRSSGFFSFESDIEHILIS 275

Db 175 -WEAPTQSPGAPGPRMRKSFSGPLRLSLAFRLQDGLRSSGFFSFESDIEHILIS 231

QY 276 GHNVQPTDTEENRMTLTIGQSEVYLISPDTKKIALEKNFKEISFCQSGIRHVDHFGFI 335

Db 232 GHNVQPTDTEENRMTLTIGQSEVYLISPDTKKIALEKNFKEISFCQSGIRHVDHFGFI 291

QY 336 CRESSGG--GGFHVCVYFQCTNEALVDEIMMTLKQAFVAAVQQTAKAPALCEGCPQLQ 393

Db 292 CRECSGGGGGFHVCVYFQCTNEALVDEIMMTLKQAFVAAVQQTAKAPALCEGCPQLQ 351

QY 394 SLHKLCERIEGMSSKTKLEQLQHLTLTNOEQATIFEVQKLRPRNEQENELIISFLR 453

Db 352 GLHKLCERIEGMSSKTKLEQLQHLTLTNOEQATIFEVQKLRPRNEQENELIISFLR 411

QY 454 CLYEKQKQKHIHIGENKQTSQMAAENIGSELPPSATRFLDMLKNAKAKSLTESLESILS 513

smooth muscle myos
micl homolog - fis
hypothetical prote
conserved hypothe
hypothetical prote
myosin heavy chain
probable nuclear p
centromere protein
myosin heavy chain
myosin heavy chain
364K Golgi complex
myosin heavy chain
oncogene 1 - human
oncogene 1 (tre-2
transcription fact
myosin heavy chain

ALIGNMENTS

```

Db 412 CLYEEKQHSHTGAPKQTLQVAENIGSDLPSPSRRLSLKNAKRLSTLESLS 471
QY 514 RGNKARGLOEHSISVDLSSLSSTLSNTSKPSVCEKALPISSEFKLLGSSEDLSSDS 573
Db 472 RGNKARGLOEHSASVDLSSSTLSNTSKLSGMDKEAFVSETSFKLLGSSDDLSSDS 531
QY 574 ESHLPEEPAPLSPQAFRRANTLSHFPLECOEPPQAPARGSPGVQKMLRYHSVSTETP 633
Db 532 EGHIAEASALLSPQAFRRANTLSHFVPECPAPPEAQSPPGVQKMLRYHSVSTETP 591
QY 634 HRKDFESKANHLGSGGTPVKTTRHRSWQIIFLAVATPQKACDSSRYEDYSELGELPP 693
Db 592 HRKDFESKANHLGTDCTPVKTTRHRSWQIIFLAVATPQKACDSSRYEDYSELGELPP 651
QY 694 RSLPEVCEDEGPPGPPPEKKTRELKWKALQIILLRMKEKNQKLOASENDLLN 753
Db 652 RSLPEVCEDEGPPGYRKGRHRHASFESCGKRPSCSRSLYRMKEKNQKLOASENDLLN 711
QY 754 KRLKLDYEEITPCLKEVTVVMEKMLSTPGRSKIKFDMKMHSAVGQVPRHHRGEIWKFL 813
Db 712 KRLKLDYEEITPCLKEVTVVMEKMLSTPGRSKIKFDMKMHSAVGQVPRHHRGEIWKFL 771
QY 814 ABQFHLKHQFPKQKQDPVYKELKQLTSQOHAILEDLGRFTTPHYFSAQLAGQLSL 873
Db 772 ABQFHLKHQFPKQKQDPVYKELKQLTSQOHAILEDLGRFTTPHYFSAQLAGQLSL 831
QY 874 YNLKAYSLLDOEVCYCGLSFVAGILLHMSSEAFKMLFMDMGLRQYRPMIIL 933
Db 832 YNLKAYSLLDOEVCYCGLSFVAGILLHMSSEAFKMLFMDMGLRQYRPMIIL 891
QY 934 QIOMQLSRLLHDYRDLYNHLHEEIGPSLYAAPFWLTMFASQPLGFVARVDMIFLQ 993
Db 892 QIOMQLSRLLHDYRDLYNHLHEEIGPSLYAAPFWLTMFASQPLGFVARVDMIFLQ 951
QY 994 GTEVIFKVALSLGSHKFLILOHLENLETIVDFIKSTPLNGLVQMEKTINQVFMEDIAK 1053
Db 952 GTEVIFKVALSLGSHKFLILOHLENLETIVDFIKSTPLNGLVQMEKTINQVFMEDIAK 1011
QY 1054 LOAYEVEHVLOEELIDSSPLSDNORMKLEKTNLSLQKNDLLEQLQVANGRIOSLEA 1113
Db 1012 LOAYEVEHVLOEELIDSSPLSDNORMKLEKTNLSLQKNDLLEQLQVANGRIOSLEA 1071
QY 1114 TIEKLSSSEKLLQKMLTLELER 1136
Db 1072 TVEKLLTSEKLLQKMLTLEVER 1094

RESULT 2
T00261
hypothetical protein KIAA0603 - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00261
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545
A:Accession: T00261
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1299 <NAG>
A:Cross-references: EMBL:AB011175; NID:g3043729; PIDN:BAA25529.1; PID:g3043730
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0603

Query Match 46.7%; Score 2828.5; DB 2; Length 1299;
Best Local Similarity 48.3%; Pred. No. 8.1e-150;
Matches 621; Conservative 184; Mismatches 291; Indels 191; Gaps 28;
QY 21 FGLQIVGLSPVHSLTMTMPLPWVAEVRRLSRQSTR---KEPVTKVRLCVSPSLRCEP 77

```

```

Db 36 FRLWVGSGSLDHRITLPLPMLPMAEIRRSQKPPAGCGGAPAAAREVILVLSAPFLRCVP 95
QY 78 EPG-----RSQWMDPLIYSSIFECKPQVRVHKLHNSHDPSPYACLIK--EDAVHRQ 126
Db 96 APGAGAGSGTSPSATQPNPAVF--IFEHKAQHSIRFIHNSHDLTYFAVLIKAAQPDPEQ 153
QY 127 SICVYFKADDQTKVPEIILSSIRQAGKIARQBELHCPSEFDOTF--SKKFEVLFCGRVTV 184
Db 154 MACHVFRATDSQVDPVIVSSIRQLSKAAWKEDAKPSKONEDAFYNSQKFEVLYCQKVT 213
QY 185 HKKAPALIDECIEKFN-----HVSG-SRGSE-----SPRENPPHAA-- 220
Db 214 HKKAPSSLLDDCMKFSLHEQRLKIQEQRGPDGDLADLEVVVVPGSGDCLPEADG 273
QY 221 -----PTGSOEVPVRRPMKSFSPQGLRSLAFRELQDGLRSGFPSS----- 263
Db 274 TDTHLGLPAGASQAPALTSRVCFPE-----RILEDSGFDEQOEFRSRCSSVTGVOR 324
QY 264 -----FEESDIENHLISGHNIQPTDIBENRTMLFTIQGSEVYLISPDTKKIALEKNFE 318
Db 325 RVHEGSKSQPRRRHASAPSHVQPSDSKNTMLFQVGRFEINLISPDTKSVVLEKNFKD 384
QY 319 ISFCSQGIHVDHFGFICRESSGGGFGFVYCVFCTNEALVDEIMMTLKQAFVAAVOQ 378
Db 385 ISSQGIKHVDHFGFICRESSPEGLSQYICVVFQCASESLVDEVMTLKQAFSTAAALQ 444
QY 379 TAKAPALCEGCPLOSLLKLCERIEGMNSKTKLEQKHLTLTNOEOATFEEOVKLRP 438
Db 445 SAKTOIKLCEACPMHSLKLCERIEGLYPPRAKLVQIRHLSLTNOEQADIFERVOKMPP 504
QY 439 RNEORENELIISFRLCLYEENKQEHIGEMKQTSQMAENIGSELPSPSAT---RFRIDM 495
Db 505 VSDQENELVILHLQCEAKQKTHVHIGEGPSTIS-----NSTIPENATSSGREFKLDI 558
QY 496 LKNKAKRLSTESLEISLRS--NKARGLOEHSISVDLSLSTLNTSKEPSVCCEKALP 554
Db 559 LKNKAKRLSTSLSEIFSRGANMRRLGSLGSDVSFERSNSLAS-----EKDYSP 606
QY 555 ISESEFKLLGSS-----EDLSSDSESHLPEPAPLSQQAQFRRRANTLSHP--I 602
Db 607 -----GSPGPTPPASPSSSAWQTFPEEDSD-SQ--FRRRAHTSHPPSSTKRKL 654
QY 603 ECQEPPOPARSPGVSQ-----RKLMYRHSYST-----ANHLGDSGGTPVKTTRHSHRQ 663
Db 655 NLQDGRAQVRSPLLRQSSEQCSNLSSVRRMYKESNSSSLPSLSTSFSAFSTAPSL 714
QY 631 -----ETPHERKDFESK-----RKLMYRHSYST-----ANHLGDSGGTPVKTTRHSHRQ 663
Db 715 KSFYQNSGRLSPQYENIRQDTASESDGEGKRTSTCSNESLSVGGTSTVTPRISMRQ 774
QY 664 OIFLRVATPQKACDSSRYEDYSELGELPPRSPLEPVCEDGPF-----GPPPEEKRR 715
Db 775 RIFLRVAFPMKNSPSAMQOQDGLDRNELLPLSPLTMEEPVIFLSEDDPEKIEERK 834
QY 716 TSRELRLEWKAQILOOILLRMKEKNQKLO--ASENDLKNRKLKLDYEEITPCLKEVTV 774
Db 835 KSKELRLWRKRAIHQOILLRMKEKNQKLEASRDELQSKRYKLDYEEYGACQKEVLTW 894
QY 775 EKMLSTPGRSKIKFDMKMHSAVGQVPRHHRGEITWKFLAEQFHLKHQFPSPKQKDPVY 834
Db 895 KKKLLN--CRAKIRCDNEDITHLLKGVFPRSRGRTWQFLAQYRLRHLRPNKQPPDISY 953
QY 835 KELLQKLSQHAILEDLGRFTTPHYFSAQLAGQLSLYNLKAYSLLDQEVGCGQLS 894
Db 954 KELLQKLSQHAILEDLGRFTTPHYFSAQLAGQLSLYNLKAYSLLDQEVGCGQLS 1013
QY 895 FVAGILLHMSSEAFKMLFMDMGLRQYRPMIILQIOMYQSLRLLHHDYHRLYNH 954
Db 1014 FVAGILLHMSSEAFKMLFMDMGLRQYRPMIILQIOMYQSLRLLHHDYHRLYNH 1073
QY 955 LEEHIGSLYAAPFLMFAFQPLGFVARVDMIFLQGTVEIFKVALSLGSHKPLIL 1014

```

Db 1074 LEENISPSLYAAPWFLTLFASQSLGFVARVEDIIFLQGTVEIFKVALSLSSQETLIM 1133
Qy 1015 QHENLETIVDFKSTPLNGLGVQMEKTIQNVFEMDIQAQAYEYEHVLQBELIDSS-P 1073
Db 1134 ECESFENIVEFLKNTLPDMNTSEMKIITQVFMEDISKOLHAYEYEHVLQDELSEYS 1193
Qy 1074 LSDNORMDKLEKTNSSLRKQNDLLEQLQVANGRIQSLQLEATIEKLLSSEKILQAMLE 1133
Db 1194 CEDSETLEKLRANSOLKQNDLLEKLOVAHTKIQALESNLENLLTRETCKMSLIRTE 1253
Qy 1134 LERSALLQTVBELRRRSAPSDREPC 1160
Db 1254 QEKMAQKTVQELRK--LLPADALANC 1278

RESULT 3
T13718
pollux gene protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13718
R:Xu, X.Z.; Wes, P.D.; Chen, H.; Li, H.S.; Yu, M.; Morgan, S.; Liu, Y.; Montell, C.
J. Biol. Chem. 273, 31297, 1998
A:Title: Retinal targets for calmodulin include proteins implicated in synaptic transmis
A:Reference number: 217709; MUID:99030403
A:Accession: T13718
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1379 <XU>
A:Cross-references: EMBL:Y17919; NID:g3893102; PIDN:CAA76939.1; PID:g3893103
C:Genetics:
A:Gene: pollux
A:Cross-references: FlyBase:FBgn0004879

Query Match 24.2%; Score 1462.5; DB 2; Length 1379;
Best Local Similarity 32.1%; Pred. No. 1.6e-73;
Matches 366; Conservative 174; Mismatches 322; Indels 277; Gaps 30;

Qy 156 QEELHCPSEFDDTSKKEVFLFCGRVTVAAHKAPALIDECIEKFN----- 201
Db 100 QSNLSLLADISPNTHFFEVAVYVGRKRVSKVPNTFIDDALPKFKAYDAQRLLRLQNRK 159
Qy 202 -----HVSGRGS 209

Db 160 NLSSEGGVGIEAKPSSSLKSHDLKEDEEQQKHGDDSDQSAQPLVQLTGAEEG 219
Qy 210 ESPRPNPHAAPTGSOEVRPRMRKFSQGLRSLAPFKELQDGLRSGSGFFSFEESDI 269
Db 220 AANRPLENKE---NKSPEKAPLRGQSQIELG-----HKEHSDGSQPSAA----- 262
Qy 270 ENHLISGHNIV---OPTD----- 284
Db 263 -NSQLEAPNVVKNQPTPPRQGVGTGTASAGPSQLHPNVAAMDNIQKDRSASQGI 321
Qy 285 ---TEENRMTLFTGQSEVYLISPDTKKIALEKNFELSFCSQIRHVHDHFGFCRESSG 341
Db 322 PPYVEQNTWYFLVGRCDRLRISDPKQVLLYKDFKVASCVHGQKSLDHFGIICRELNN 381
Qy 342 GGGFHFVYVFOCTNEALVDEIMTTLKQFTVAQAQAKAPAL--CEGCPLOSRLKLC 399
Db 382 DG---YIGYVFKQSEHVCDDIVAIAQAQFTCA--EQKKKQDTQIFSCHECPMLWYHKLC 437
Qy 400 ERIEGMNSKTKLELQKHLTTLTNOEQATIFEVQKLRPRNE--ORENELIISFLRCIYE 457
Db 438 TDVEGLSEKKTQALILRIETLSDDQEBIVWAKFCGSEKTNPSVAEQNFLMMLLRAHCE 497
Qy 458 EKQEHIIHIGEMKQTSQAAENIGSELPPSATRFLDMLKNKARSLTESLESILSRGNK 517
Db 498 SRQQRHVH-----DTAENRSEFLNQLGGST-----IFMKAKRSLTNSFDNLKR--- 542
Qy 518 ARGLOEHSISVDLSSLSLTSNTSKPSVCEKALPISSESKLL--GSSEDLSSDSES 575

Db 543 -----KPS-----KDIIVPSHNLDRIDREGSABPLGTQ----- 570
Qy 576 HLPPEAPLSPQQAARRRANTLSHFPIECQEPQPARSGPGVSQRK--LMRYHSYSTETP 633
Db 571 -----SPEGFRSRNTVGASP-----SSKPTAEQLKSPMMDIFIKVGNP 611
Qy 634 HERKDFESKANHLGDSGGTPVKTTRRSHRQOIFLRVATPQKACDSSSYEDYSELGELPP 693
Db 612 KE-----AETHQGSWRQAILNSVVTTPSKGLDSEVTEFLS----- 646
Qy 694 RSPLEPVCEDGPGFPPEKRTSRELRLWOKAILQOILLRMEKENOKLQASENDLIN 753
Db 647 -----PMKPAKRGKRDAAELRELWRTAIRTQILNRMETENALQARQENEL 695
Qy 754 KRLKLDYEEITPCLKEVTVVMEKML---STP-GRSKIKFDMKMSAVGGQVPRHRHGEI 809
Db 696 KRIKLDYEEIVPCDKQLIERWEQIERNSTQIGNKK---DPKVLGHAIRTVGPRSKRGDV 752
Qy 810 WKFLAEQFLHAKHQP--SKQPK-DVPYKELLKQLTSQOHAILEDLGRFTPTHPYFSAQL 866
Db 753 WTFLAEQ-HSMNTAPVDTKRFPNFTPYHMLLKHLEHQHAFIDLGRTPFNHQFYKDP 811
Qy 867 GAGQLSLNLIKAYSLLDQEVGYCOGLSFVAGILLHMHSEEAFAKMLKFLMDFMDGLRKQY 926
Db 812 GLGQLSLENLLKAYSILDPGLCYGCGLFCVGLLLHCDANSFQLLKHLMFRNRRTKY 871
Qy 927 RPDMLIIOIOMYQSLRLHDYRDLNHLBEHIGPSLYAAPFWLTMFASOPLGFVARV 986
Db 872 LPDKKFKQLYQLSRLVKDHLPLVWLDQNDVSPTLYAAPWILTVFSOPLGFVARV 931
Qy 987 FDMFLQGTVEIFKVALSILGSHKPLIIOHENLEITVDIKSTPLNLGVOMEKTIQNOVF 1046
Db 932 FDLFLLESSDVIFKPAIALLSVHKQQLAKONFEIMDYLTVTVPKMEHTCMEQIMKLVF 991
Qy 1047 EMDLAKQAYEVEYHVLQEEELIDSSPLSDNORMDKL--EKTNSLRKQNDLLEQLQVA 1104
Db 992 SMDIGKQLAEVNEVNLQEEI-----TTNHHLEMLNREKT-----QNOHLEQLQQA 1040
Qy 1105 NGRIOSLEATIEKLLSSSEKILQAMLTLELRSALLQVVEELRRRSABSPDREPTOP 1163
Db 1041 QSSIAQLETT-----RSSQQAQITTLQSQVQSLELATIQTILGRYVGVOLVHNPDLELP 1092

RESULT 4
T13163
Rab6 GTPase activating protein, GAPCenA - human
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13163; T08800
R:Culfi, W.H.; Possmayer, F.; Zander, H.; Bordes, N.; Jollivet, F.; Couedel-Courteille
EMBO J. 18, 1772-1782, 1999
A:Title: Characterization of GAPCenA, a GTPase activating protein for Rab6, part of v
A:Reference number: 217619; MUID:99219856
A:Accession: T13163
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1030 <CU>
A:Cross-references: EMBL:AJ011679; NID:g4582148; PIDN:CAB40267.1; PID:g4582149
R:Ansorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: 216472
A:Accession: T08800
A:Molecule type: mRNA
A:Residues: 785-1030 <ANS>
A:Cross-references: EMBL:AL050195
A:Experimental source: adult uterus; clone DKFZp586D2123
C:Genetics:
A:Note: DKFZp586D2123.1

Query Match 9.6%; Score 580; DB 2; Length 1030;
Best Local Similarity 23.9%; Pred. No. 1.8e-24;
Matches 221; Conservative 137; Mismatches 311; Indels 254; Gaps 33;

QY	295	IGOSEVYLISPDTKKIALEKNEKEISFCQSQIRHVDHFG-----FTCRSSSGGGGFHFV	348
Db	152	VSGEIVRLDPDQNTIEANIPIYKILFCVVG-----HDGTPESDCFAFTESHYNAEL--FR	205
QY	349	CYVFOCTNEALVDEIMMTLKQAFVAAVQQTAKAPAQLCEGCPQLQSLHKLCEREGMNSS	408
Db	206	IHVFRCEIQEAVSRILYSFATAFRSAKQTPLSATA--APQTPSDIFTVSVLE-----	258
QY	409	KTKLELQKHLTWLTNOEQATIFEVEVKILPRNREORENELIISFLRC--LYEEKQKEHIHI	466
Db	259	-----IKEDDGKGYFSAV--PKDKQRQ-----CFKLQGGIDKKIVY	293
QY	467	GEMKQTSOMAENIGSELPPSATREFRLDMLNKAKRSLTESLEILSRGNKARGLQEHSI	526
Db	294	VQOTTNKEIAIERC-----FGL-----LLSPGDKVNRNSDMHLL	326
QY	527	SVLDLSSLSLTLSNTSKEPSVCEKALPISSESKLLGSSDLSSSESHL-----PEE	580
Db	327	--DLES-----MGK-----SSDGKSVYITGWSNPKS	350
QY	581	PAPLSPOQAFFRRANTVLSHPPTECOEPPQARGSPGVSQRLMKRYHSVSTETPHERKDPE	640
Db	351	P-----HFQV-----VNEETPKDKVLFM	368
QY	641	SKANHLGDSGGTPVKTRRHSWQQIFLRVATPQKACDSSRSRYEDYSELGELPPRSLEPV	700
Db	369	TTAVDL-----VIT--EVEQPVRFLETKVRVCSNERL-----	400
QY	701	CEGPGPGPPPEKKRTSRE-----LRELWKA-----ILOQILLRMXENQKLO-----	745
Db	401	-----FWP-----FSKRSTTENFFLKQIKQIKRKNNTDLYEVVCLSESESERERKKTAS	452
QY	746	-----ASENDLLNKLKLDYEE-----ITPCLKEVTVVWEKMLSTPGRS	784
Db	453	PSVRLPQSGSQSSVISPPEDEDEEDNDPELLSGSDGVSKCAEKLKTWCELLS--KWHL	511
QY	785	KIKPDMKMHSAVGQGVPRHHGEEIKWFLAEQFHUKHQQFPKQPKQKVPYKELLKQLTQS	844
Db	512	NLAVRPKQLSSLVRNGVPEALRGEVWQLLAGCHNNDHLVEK-----YRILITKESPO	563
QY	845	QHAILDLGRTPTTPPYFSAQLGAGQLSYNTLKAYSLDQEVGVCGCLSPVAGILLHM	904
Db	564	DSAITRDINTFPADHYFDKTCGDGQDSLYKICKAYSVDYEIGYCQCGQSPFLAAYLLHM	623
QY	905	SEBEAFKMLKFLMDFMDLRKQYRPMWILQIQMWQLSRLLHDYHRDLYNHLHEEIGPSL	964
Db	624	PEEQAFSLVKIMFYDGLRELEKQNFEDLHCKFYQLERLMQVEYIPDLYNHPFLDLSLAHM	683
QY	965	YAAPWLTWPASQFPPLGIFVARYFDMIFLQGTVEVIFKVALSLLGSHKPLLOHENLEIIVD	1024
Db	684	YASQWFLTFTAKFPLVWYFHIIDLLECGISVIFNVALGILLKTKSKDOLL--TFEGALK	742
QY	1025	FIKSTPLNLGLVOMEXTINQVFE---MDIA--KQLQAYEVEYHVQLQBELI--DSSPLSDNQ	1078
Db	743	FFRVQLPK--RYSEENAKMLACNMKISSQKLKLYKEVHTMREQAQOQEDPIERFE	800
QY	1079	RMD-KLEKTNSLRKQNLDDLQOLQVANGRTQSLEATTEKLLSSESKLKQAMLTLELERS	1137
Db	801	RENRRLQEANMRLEQNDLALHEL-----VTSIALRKLDNAEKADALNKKELLMTQK	854
QY	1138	ALLQTVTEELRRRSRAPSDDREPC	1160
Db	855	KLIDABEEKRRLEESAQLKEMC	877

RESULT 5

G89453

protein F35H12.2 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001

C:Accession: G89453

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio
A/Reference number: A75000; PMID:99069613; PMID:9851916
A/Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A/Accession: G89453
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1142 <STO>
A/Cross-references: GB:chr_X; PIDN:AAA83178.1; PID:g1109865; GSPDB:GN00028; CESP:F35H12
C/Genetics:
A/Gene: F35H12.2
A/Map position: X

[illegible]

Db 739 EKEIGHQANTLRLERENDDLAHELVTLSKIELRRKLDVAEDQIETSANAIELRL-----T 792
QY 1126 QAMUTLEERSAL--LQTVLELARR 1149
Db 793 RQNDMILEENKMLREYEQIKEMYYR 818

RESULT 6

S61015
hypothetical protein YPL249c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein P1018
C:Species: Saccharomyces cerevisiae
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Nov-1997
C:Accession: S61015; S65278
R:Pohl, T.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S61010
A:Accession: S61015
A:Molecule type: DNA
A:Residues: 1-894 <POH>
A:Cross-references: EMBL:Z67751; MID:g1061234; PID:g1061240
R:Pohl, T.M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64899
A:Accession: S65278
A:Molecule type: DNA
A:Residues: 1-894 <POH>
A:Cross-references: EMBL:Z73605; MID:g1370512; PID:e246954; PID:g1370513; MIPS:YPL249c
A:Experimental source: strain S288C (AB972)
C:Genetics:
A:Map position: 16L

Query Match 7.7% Score 465; DB 2; Length 894;
Best Local Similarity 23.2%; Pred. No. 3.9e-18;
Matches 191; Conservative 153; Mismatches 292; Indels 188; Gaps 32;

QY 412 LELQKHLTLTINQQTIFEVQKLRPNQRENELIISFLRCLYEKQKEHIHIGEM-- 469
Db 116 LOVEKQDVT-----DVQAPNGLNIEKEYDAV-----KENEKYADTKVEVS 156
QY 470 -KQTSQMAENIGSELPSPATFRFLDMLKNKAKSLTESILSRGNKARGLOEHSISV 528
Db 157 SPENREVTGKNGSGE-KSSSKFLDD--ESGTTTAAANDISISSEVTPRESSENDNQI 213
QY 529 DLDSLSSTLS-NYSKESVCEKEAL---PISESSFLLGSSSEDLSSDSSEHLPPEAPL 584
Db 214 HITNEVAAGINLENKE---QKAAIEDGPV-----TAENLSSETARKVPIPTQI 260
QY 585 SPQAFRRANTLSHFPIECOEPPQPAR-----GSPGVSORKLMRYHSYSTET---PH 634
Db 261 INEKGDNSSNEVSAIP-TTSSPLPRQNVATSTSPKLPGRKQREOPPRTKNVPPPL 319
QY 635 E-----RKDFSKANHLGSGTPVKTRHRSWROQIFLURVATPQKACDSSSRYEDYS 686
Db 320 EEMKSEKFRNFE-----EYKNSYHHVPLTGSTAQ-----LESTA 357
QY 687 ELGELPPRSPLEPCVDGPGPPPEEKRTSRRELWQKAILQOIL-----LLRMEK 739
Db 358 EINLIASR-----YKTSHLNKEGETRESLQEG--QSFLKSTFTSFLENLSE 404
QY 740 ENQKLOASENDLLNKLKLDYEETPCLEKVTYTWKMLSTPGRSKIKFDMKMHSAVGQ 799
Db 405 YNEVENNEED--REMPKIDWSFTQVVDYATVASN-----EPENLEAHVTN 450
QY 800 GVPRRHGEIMKFLAEQFHLKRPSPKQPKDVPYKELLKOLTQQOHAILLDLGRTPPH 859
Db 451 GIPQIRGIWQMAN-----SKSREMEDIEYELLDTCEHLHEATIRDLART----- 497
QY 860 PYFSAQAGAGOLSYNLKAYSLDDQGVGYCGQLSFVAGILLHM--SEEEAFKMLKFLMF 918
Db 498 -KFVAE--DKMESLYKVIKYSVDPDVGYTQGMGFTAAAPLLINCNEAESFGLVLGLMK 554

QY 919 DMGLRKQYRPMIILQIQMYQLSRLLDHYHRDLYNHLEEHGSPGLYAAPWFLTFASFQF 978
Db 555 NYGLRELFLPGMGLMLMLYQFDRLLSEHSPSLNRLIRIGISSTMYATQWFLTFAYKF 614
QY 979 PLGFVARVDFMIFLQGTVEIKFVALSLGSHKPLILQHENLETIVDFTKSLPLNGLV-- 1036
Db 615 PLEFVLRIFDIVFVEGIEVLKFAVNLMKNEETLVK--LRFDELLDFLKDELFWYLAGN 673
QY 1037 QMEXTINQV-----FEMDI-----AKQQAQYEVYHVHQEE 1067
Db 674 QDDASVQVQGLSNGSNFKGNDGTFSYNVDFVHDAMGTGVITPLTLRRYRGYVEIHE- 732
QY 1068 LIDSPISDNQMDKLEKTNSSLRKQNLDLLEQLQVANGRIQSLQLEATTEKLLSSSKLKQ 1127
Db 733 -----KEQKEDHYESLRQNHQLOREAQKLEHDYSILN---KENISAANELIQ 778
QY 1128 AMLTLEL---ERSALLOTVEELRRRSAPSPRE---PECTOPE 1164
Db 779 NRLNWEMLLDEKNDLINTITDIKSQIEEIRKQNLPNPDASLPK 822

RESULT 7

T49801
hypothetical protein B11B22_30 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49801
R:Schulte, U.; Align, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakutu
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49801
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-876 <SCH>
A:Cross-references: EMBL:AL356834; GSPDB:GN00116; NCSP:B11B22.30
A:Experimental source: BAC clone B11B22; strain OR74A
C:Genetics:
A:Gene: NCSP:B11B22.30
A:Map position: 6
A:Introns: 75/3; 190/1; 449/3

Query Match 7.0% Score 421; DB 2; Length 876;
Best Local Similarity 24.0%; Pred. No. 1.1e-15;
Matches 186; Conservative 112; Mismatches 298; Indels 180; Gaps 27;

QY 421 LTNOEQATIFEEVOKLRPNQRENELIISFLRCLYEKQKEHI-HIGEMKQTSQMAEN 479
Db 133 IKEEQEDVPE---TPAQEQ-----ESKSLHRISHASKVSEASAFSA-- 171
QY 480 IGSELPPSATFRFLDMLKNKAKRSLTSLESILSRGNKARGLOEHSISVDLSSLSST-- 537
Db 172 -----ASLDVSLNDNSTIGDKTGNPEKSLDTAPSSSTKK 205
QY 538 LSNTSKEPSVCEKALPISESEF-KLLGSSSEDLSSDSSEHLPPEAPLS-----PQAFR 591
Db 206 LSISIASAL---PAMWSPSRTDEAVKSPPPAAAAAAPPAPAPALTPAPPPVTR 262
QY 592 RRANTLSHF-----PIECQEPPOPARG--SPGVSORKLMRYHSYSTETTPHERKDFESKANH 645
Db 263 KLTSPFSLWLSRSSSKEQANPPPPAATAAPRNTASSVATLSSNPDGTGLGVDEEGNKT 322
QY 646 LGDSGGTPVKTRRHSWROQIFLRY--ATPQKACDSSSRYEDYSSELGELPPRSPLEPCVD 703
Db 323 LKDR-----FKQLRREEGSAPTGDDEKSTPEEKAQV--PPSPPLPPT--- 366
QY 704 GPGPPPEEKRTSRELK-----ELWQKAILQILLRMEKENQKLOASENDL 751
Db 367 -----PNPALAPGTASGTAGPAGSDTPVDWDLWQ----- 396
QY 752 LNKRLKLDYEEITPCLKEVTYTWKMLSTPGRSKIKFDMKMHSAVGQVPRHRHGEIWK 811
Db 397 -----TVVYEGPAAVARTSP-----EQLKEAIAKGIPSAIRGVIWQ 432

QY 812 FLAEPFLKHQFPKQKDPVYKEL--LKQLTQQHAILIDLG-RFPPTHVPFSAQLGA 868
Db 433 VLQA-----SKNELETIVRDLDEKEMLRKLEKVIIRDIGART--SYSFARAAQGL 481
QY 869 GQLSLYNILKAYSLIDQEVGYCOGLSFVAGILLHLMSEEEAFKMLKFLFMDGLRKOYRP 928
Db 482 -QEGFGVCAYALFDEAVGAYQGMFLVMPPLFNPMPEEAFCLLVRLMNYHLRELFIQ 540
QY 929 DMILIQMTQLSRLLHDYRDLNYHLEHEIGPSLYAAPFWLTMFASQPPLGFVARVFD 988
Db 541 DMPGLHKCLYQFERILEDELPALYCHLHRRGISPHLYATQFMTLFAIRPPLQLVRLIYD 600
QY 989 MIFLQTEVFVKVALSLGLSHKPLIHOHENLETIVDFIKTFLNLGLVQMEKINOVFEM 1048
Db 601 LIFEGLSAILKIGCIVLMKNATALLGMSDMSQLTTFKDRLEFDV-YIDATPSSNILEN 659
QY 1049 DIAQQAQVEYVHVQEEILDSPLSDNORMDKL-----EKNSSLRKQNLDLLEQ 1100
Db 660 GFFGSSA-SIDKEVYRADQVLRDADCVNITAEALLKAYGKEWEEKTKA--EKEREQLEG 716
QY 1101 LOVANGR---IOSLEATIEKLSSSKLQKQAMLTLELERSALLOIYVEELRRSAE 1152
Db 717 LKQANTNYTVVRRLEERIEAVDREQASLATELVRTKVENELRDENESLRGQVKE 772

RESULT 8

T29570

hypothetical protein T24D11.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29570

R:Wu, X.; Gattung, S.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid T24D11.

A:Reference number: Z20643

A:Accession: T29570

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-480 <WUX>

A:Cross-references: EMBL:U49940; PIDN:AAA93411.1; CESP:T24D11.1

C:Genetics:

A:Gene: CESP:T24D11.1

A:Introns: 59/1; 91/2; 123/3; 192/3

Query Match 5.6%; Score 397; DB 2; Length 480;
Best Local Similarity 25.7%; Pred. No. 9.9e-15;
Matches 123; Conservative 83; Mismatches 158; Indels 114; Gaps 15;

QY 661 WRQOI-----FLRVATPQKADSSSRDYDS---ELGELPPRSPLEPCEDGPGFP 708
Db 26 WKRLLETSPPHFFPAPATTTACIFGSKLVAGSADTECGE-----PCETG---- 72
QY 709 PPEKRTSRELRLMOKAILQOI-LLLRMKEKNQKLAQENDLLNKLRL----- 756
Db 73 -----APVSLNEVDLLAKMEQLNKSNEEDSRVASKTSSSRKRGARE 116
QY 757 ---KLYEETPCLEKVTTWKEMLSLTPGSRKTKFKDMKMHSAVGQVPRHHRGETWKFL 813
Db 117 HSPDEEDLWSYNGELILNWE-----IEVKRPNYIKDLVKRGIPQHFRIAMQNL 168
QY 814 AEQFHLKHQFPKQKDPVYKELKLTQQHAILIDIGRTPTTPYFSAQLGAGQSL 873
Db 169 SN-----ASVSSVHDL-YSDMKQSSVYEVKIQIDIPRTYPELDFFK-DGERQSL 218
QY 874 YNLIKAYSLDQEVGYCQGLSFYAGILLHLMSEEEAFKMLKFLFMDGLRKOYRPMIIL 933
Db 219 FNVIKAYSVHDKVGYCQGSFAFVGLLLQMPPEEAFVLVSLMENVRLRYLKYKPTMTDL 278
QY 934 QIQMYQLSRLLHDYRDLNYHLEHEIGPSLYAAPFWLTMFASQPLGFVARVFDIFLQ 993
Db 279 GLCMFQLECLVQDQMPDLXYTHFNMMGFDTSKYASSWFLTLFTTTMTPLDIANRIMDCFLVE 338

QY 994 GTEVIEFKVALSLGSHKPLIHOHENLETIVDFIKSTLPNLGLVQ-MEKTINQVFEKD--- 1049
Db 339 GMDFFICISIA-----ILQARIELL-----RLDMGMLKYFOREVREYEFADL 384
QY 1050 ---IAKQLQAQYVEYHVLAQEEILDSPLSDNORMDKLEKTNSSLRKQNLDLQLQVA 1104
Db 385 LFTVANQVQL-----NAKRMKLEKDYLTFRKTKQEQAVALRVS 423

RESULT 9
T38411
probable GTPase activator protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38411
R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21792
A:Accession: T38411
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1031 <BRO>
A:Cross-references: EMBL:Z73100; PIDN:CAA97366.1; GSPDB:GN00066; SPDB:SPAC26F1.09
A:Experimental source: strain 972h; cosmid c36F1
C:Genetics:
A:Gene: SPDB:SPAC26F1.09
A:Map position: 1

Query Match 6.4%; Score 387; DB 2; Length 1031;
Best Local Similarity 20.7%; Pred. No. 1.1e-13;
Matches 207; Conservative 153; Mismatches 377; Indels 262; Gaps 39;

QY 206 SRGSESPRPNPAAPTGSOEVPVRPMKRSFSQPLG-RLSLAFKELQDGLRSGSGFFSF 264
Db 144 SESTEFPVNEVNETATIGNEDVGE--RSGFPSGEGLDNEPESQORDLDGTG-----NL 192
QY 265 EESDIENHLISGHNIQVQTDIENRTMLFTIGQSVLYLISPDTKKIALE-----KNFKEI 319
Db 193 APEDLKDEKSVHEFNEPNDLRQ-----QEEYSDDDDINVEFEDVNEIENHQL 243
QY 320 SFC--SQGIRHV-DHFGFTCRESSGGGFHFVYVQCTNEALVDEIMTFLKQAFVAAV 376
Db 244 SVADEDTSLRVKGMIFVGKEDFGEA-----DISNSVFEIQ---NGPNSDITVSGF 292
QY 377 QQTAKAPALCECPQLSLHKLCLEREGMNSSKTKLEKHLTLTNOEQATIFEVQKL 436
Db 293 KETSSI-----VNSSST-----TEKPGVALDSQNDTSIF----- 321
QY 437 RPRNEQRENELIISFLRLCLYEEKQKEHIIGEMKQTSOMAAENIGSELPPSATRFRDML 496
Db 322 ---NEQTSNLSLTF-----NDLTLDHL-----PENVESE--PVACK----- 353
QY 497 KNAKAKSLTESLISLRGNKARGLQHSISVDLSLSLSTLSTNTSKEPSVCEKE----- 551
Db 354 -----ENETAKNESGASDNDHKANVHFVFLKSEDAITLNEEKIATQDDPLEA 401
QY 552 ALPISSESKLLGSSDLSSESHPLEPAPLSPOQAFRRANTLSHPPIEC-----OE 606
Db 402 PTPIVASSSTIFUNSNORNDLSGASQEPHP-----KDGNTNSSLPLDNTNNLSNSE 454
QY 607 PPQAPAGSPGVSRQKMLRVHSVSTETPHERKDPESKANHLGDSGGTP-----YKRRHSW 661
Db 455 PPSHVLDASETIEVTQIKKLQNVQPETIKDEVGKN-TAFSPGTSLSNHHVTKSRSA 513
QY 662 ROQIFURVATPKQACSSSRDYDSYSELGELPPR---SPLEPVCEDEGFPFGPPPEKRTSR 718
Db 514 HNNSTSPFSTAVSSWLNPLRYP-----SDKSPRVISYLESVFISKP-----R 556
QY 719 ELRELQAKLLOQIILLRMKEKNQKLAQENDLLNKLRLKLDYEEITPCLKEV----- 770
Db 557 SIGDAOKLEILEYLOQSOSTSVNQVFTLLSNFTQNPLFVLDD-----ECDFENILMHN 611

QY 771 -----TTVWEKMLSTPGRSKIKFDMKMH-----AVCGVPRHRHGRGIWFLA 814
Db 612 SHTVHTVWKTI-----SSWTSYDYMOSYSLKNCSDSKAIRKDLDRTPAPEI----- 661
QY 815 EQFLHKHOPPSKQO--PKDVPYKELLKQLTQQHAILDLGRTPPTHTPYFSAQAGAGOLS 872
Db 662 ----LSHFFSNRQLEPTD-----NIAESTAN 684
QY 873 LYNILKAYSLLDQEVGCGQLSFVAGILLHMSSEAFKMLKFLMDFMGLRKQYRPMII 932
Db 685 LHRVLSLAIVLPQVGTQGSWIGALLMHLPAQAFALLVFLKNVHLQNIFFSSEMRG 744
QY 933 LQIQWQLSRLLHDYHRDLNHLHEEIGPSLYAAPFLTMFAQFPLGFMIFL 992
Db 745 LSRVLHQPTRLVEDYMPSLATHFRKQDIKTCYASEMFLTLFAYKFLPVEVVAHYLDLFL 804
QY 993 QGTEVIFKVALSGHKLPLJLQHENLTIIVDFIKTLPLNLGLVQMEKT----- 1041
Db 805 YPGILFNGALL-SHSQESLLKLNMDRLISYLKEDI-----FLAFKETQGEYNDTSLF 859
QY 1042 INQVFEMDIKAO-QOAYEVEYHVL--QBELIDSSPLSDNORMDKLEKTNSSLRKONL--- 1095
Db 860 VKTAFSEFIQPDVLDRYGNEYDILLKSEHLDSS-----LEEMNRNKHSLNEHFIMLS 912
QY 1096 DLLEQLQVANGRIOSLEATIEKLSSSEKLAQAMLTLEL 1134
Db 913 DSMANLQVHEHNMAL-LLKERMVYLNKQTVQASLKSEI 950

RESULT 10

T51376
plant adhesion molecule 1 (PAM1) - Arabidopsis thaliana
N:Alternate names: protein FIN13_70
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Nov-2000
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Ban
submitted to the Protein Sequence Database, August 2000
A:Reference number: 225393
A:Accession: T51376
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <SAT>
A:Cross-references: EMBL:AL391145
A:Experimental source: cultivar Columbia; BAC clone FIN13
R:Stevenson, B.A.; Zhu, J.K.
submitted to the EMBL Data Library, June 1998
A:Reference number: 225910
A:Accession: T52031
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-356 <STE>
A:Cross-references: EMBL:AF069528; PIDN:AAC33763.1
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 5
A:Introns: 48/2; 113/3; 170/3; 209/3; 275/3; 328/3
A:Note: FIN13_70

Query Match 6.48; Score 386; DB 2; Length 356;
Best Local Similarity 31.4%; Pred. No. 2.6e-14;
Matches 85; Conservative 63; Mismatches 107; Indels 16; Gaps 4;

QY 742 OKLAQSEDLNKLKLDYEITPCLEVTWVWEKMLSTPGRSKIKFDMKMH---SAVG 798
Db 28 QEHSSSPORFTKTSINYEK---EKRVTYKWKRMIGTGGSDWKHYVRPYPVVKRRIR 83
QY 799 QGVPRHRRHGWKFLAEQFLHKHOPPSKQKDPYKELLKQLTQQHAILDLGRTPPT 858
Db 84 KGIPDCLRGLWQLISGRDLLANPG-----YVQLVIYVETSASELDIIRDSRTPS 137

QY 859 HPYESAQLGAGQLSLYNILKAYSLLDQEVGCGQLSFVAGILLHMSSEAFKMLKFLM- 917
Db 138 HVFTQKRHGPQGRSLYNVLRKAYSYVDRGVYVQGMGTAGLLLYMSEDAFWLLVALLK 197
QY 918 --PDMGLRKQYRPMIILQIQWQLSRLLHDYHRDLNHLHEEIGPSLYAAPFLTMFA 975
Db 198 GAVHSPTEGLYQAGLPLVQVQLLQFDQLVRLMPLKGLGHEFTQEMINFSWYASQWFIIVFS 257
QY 976 SQFPLGFVARVDFMIFLQGTVEVIFKVALSLL 1006
Db 258 YSLFFHSALRIWDVFLAEGVKIVFKVGLALL 288

RESULT 11

S62481
hypothetical protein SPAC4G8.04 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 31-Jan-2000
C:Accession: T38849; S62481
R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z21745
A:Accession: T38849
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-772 <BA2>
A:Cross-references: EMBL:Z56276; NID:g1022345; PIDN:CAA91205.1; PID:g1022349; GSPDB:g
A:Experimental source: strain 972h-; cosmid c4G8
C:Genetics:
A:Gene: SPAC4G8.04
A:Map position: 1L
A:Introns: 680/3

Query Match 5.78; Score 343; DB 2; Length 772;
Best Local Similarity 22.18; Pred. No. 2e-11;
Matches 198; Conservative 126; Mismatches 340; Indels 232; Gaps 34;

QY 294 TIGSEVYLISPDPKTKIALEKNFREISFCSGIRHVDHFGICRESSGGGFFHVCYVQ 353
Db 4 SVSDSSVYLSAPTAKPIA-----SGSVATG----- 28
QY 354 CTNEALYDEIMMTLKQAFVAAVOOTAKAPALCEGCPLOSLKLCERIEGMNSKTKLE 413
Db 29 -----IDS--MPSKVDITPCDLLNKSAPLFVECNQESLHIPSGLHVPASIERL 80
QY 414 LQKH-LTTLTNEQATIFEE---VQKLPRNEQRENELIISFLRCLYEKQKH---IHI 466
Db 81 IEKHGAVNLLRQLAKDVAERDSFISDLKFHESR-----YVRELLREHGLDPLV 131
QY 467 GEMQTSQMAAENIGSELPPSATR-----FRDLMLNK---AKRSILTESLE 509
Db 132 ANTKLSQRHSASFPPSSQEPSPENPSLTGKPHLYARIDSAINEPTPSDRSLSPSLV 191
QY 510 SILSRGNKARGLOEHSISVDLSDSLSTLSNTSKEPSYCEK-EALPISE-----SS 559
Db 192 PLL---KLPAI-DHAYSSSSSSDLPSDPNSASYIASSKQKASSLKLTSLKFFYSWSS 246
QY 560 FKLGGSEDLSSDESHLPEEPAPLSPQAFRRANTLSHPPIECQBPQPARGSPGVQ 619
Db 247 SSLQHTRENLDHSTSLRDHDPSSLSSSKPPRSSP-----RCSTPSVSTFVSATSE 298
QY 620 RKLMYHSVSTETPHERKDFESKANHLGDSG----- 651
Db 299 PEVETY-SVSTKNSSSNKLRSLSKLLSTNLNKNKPLSLSTAPSMPISIGSELGNMIP 357
QY 652 --TPVKTRRSRQOIFLRVATPQACDSSSRYEDYSELGELPPR-SPLEPVCEGDPFGP 708
Db 358 KETOPPSNRNDWKDYL-----DNNSK-EILOQFGFLQKRPSTDTPLC----- 398
QY 709 PPES-----KKRT-----SRELWQKAIL---QOILLRMEKNQKLOA 746
Db 709 PPES-----KKRT-----SRELWQKAIL---QOILLRMEKNQKLOA 746

Db 399 -PEDIKLNQKOTLSFYESNYGLVDDFFGNELDGLNDSPILLNKKDILLDMKESTRQKNWS 457
QY 747 SENDLLNKKLDYBEITPCLEKVTWVWVKMLSTGRSKIKFDMKMHSAVQGVPRHR 806
Db 458 LFFORLYKKYKT-DEDTIGLIGISSIGVK-----GRHG-KRWKHFRELKNGVPLCYK 510
QY 807 GEIWKFLABQFHLKHQFPSPKQPKDVP--YKELLKQTSQQA-----ILIDLGRTFPTHP 860
Db 511 AKVWLECSGAYQL-HS-----PGYVELLSRTDEVSASVAQIDMDINRTMAKNV 559
QY 861 YPSAQLGAGQLSYNLKAYSLDQEVGYCQGLSFV-AGILLHMSSEAFKMLKFLMPD 919
Db 560 FFGGK-GPGIPKRLRLVAYSRRHNPHIGYCOGMNIGAFLLLYLASEDAFYMLMSIEN 618
QY 920 MGLRKOYRDMILQIOMYQLSRLLDYHRLYHLEHEIGPSLYAAPWFLTMASQFP 979
Db 619 VLPPKYFTPLMTSRADQLVLSFKVESLPEIYSHLELGLDLDALISFWLVSFTDTLP 678
QY 980 LGFVARVDMIFLOQTEVIFKVALSLLGSHKPLIIOHENLETIVDFIKSTPLNGLVOME 1039
Db 679 TNISFRIFDMLFCGVCVLFRAVATILKSLKQILACNSSSAIYSFLSDL-----728
QY 1040 KTNQVFMEDIAKQLOAYEVEYHVLQBELIDSSPLSDNORMDKLEKTNSSLRKQNL 1095
Db 729 -----VOYSFQDPSFIKEAA-----DRWSKLVTEKSIERKRL 761

RESULT 12

S50920
hypothetical protein YMR192w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR192w (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S50920
R:Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A:Reference number: S50917
A:Accession: S50920
A:Molecule type: DNA
A:Residues: 1-720 <PEA>
A:Cross-references: EMBL:Z47815; NID:G642280; PIDN:CAA87813.1; PID:G642284; MIPS:YMR192w
C:Genetics:
A:Map position: 13R

Query Match 5.2%; Score 317.5; DB 2; Length 720;
Best Local Similarity 21.0%; Pred. No. 4.9e-10;
Matches 149; Conservative 126; Mismatches 24; Indels 191; Gaps 26;

QY 557 ESSFKLLGSSEDLSS--DSSEHLPEE-----PAPLSPOQAFRRRANTL--SHFTPECOE 606
Db 92 ESQSKTILPSDDLSQLQLETEESKVEEALAKRITSPPLP-----RADCTEESASAKSSL 145
QY 607 PP-----QPARGSPGVSKRLMYHVSVTETPHERKDFESKANHLGDSGGTPVKRR 658
Db 146 PPVLGNKNDQAPLDLPQPLPQVNAETLHLKAPHGN-----183
QY 659 HSWRQOIFLRVATPQKACSSRYEDYSELGLPVR--SPLEPCVEDGPFPPPEKKR 715
Db 184 -----ATPSKSPTSVAGNSSSTPTPLPPRIEDLDLAAQ-----219
QY 716 TSRELRLWQKAILQOILLRMEKENQKLOASENDLLNKLRLDYEEI--TPCLREVTVV 773
Db 220 -----KHFLASTFKRNLKYKSEDSIKCDLKN-ILN--LKEDSKINNNEIPEVSVS 271
QY 774 WEKMLSTPGRSIKTFDMKMHSAVQGVPRHRHGETWKEFLAQFHLKHQFPSPKQPKDVP 833
Db 272 WLKVIQDYQNILN-DIETLHQLSGRIPAAVRLVWQLVSY-----AKRSFDFI 321
QY 834 YKELLKQTSQQAIIIDLGRTFPTHPYPSAQLGAGQLSY-----NILKAYS 881
Db 322 YETYLTEMAP-----FDVQEFNQLKMDVEVPSEYKRVISNVLKAYL 363

QY 882 LLDQEVGYCQGLSFVAGILLHMSSEE--AFKMLKFLMDFMGLRKOYRDPDMLIIOQMVO 939
Db 364 LFDPECFESTDIAYIIN-MILDVCEBEANAFGLLRLKMKVYGLRLFLPASEIDILCYK 422
QY 940 LSRLLHDYHRLYHLEHEIGPSLYAAPWFLTMASQFPGLFVARVDMIFLOQTEVIF 999
Db 423 FDRLEVEEYPEIHNHMKVGRSSMFLPGFFTTLFOKKLPTETIQRIGMDMVEFLGDSIM 482
QY 1000 KVALSLLGSHKPLIIOHENLETIVDFIKSTPLNGLVOMEKT-----INQVF 1046
Db 483 RILATLLSNSRDHLK-WGFDDMLLELLAKSGLLDAYIKQNDGTRGDTLLSNECDMLLQDS 541
QY 1047 EM DIA---KQLOAYEVEYHVL-----QBELIDSSPLSDNORMDKLEKTNSSLRKQNL 1094
Db 542 MMKVAITPKTMKKYSEVEIHRLDNEKEVQYKSTERNLHQLKHKVRKLENDYTSLNREH 601
QY 1095 LDLLSQ-----LQVANG-RIQSLEATIEKLLSESK-----LQK 1127
Db 602 VTIANELVKRLNIEVNLNENNGYKLIQILD--LRKKLDSEKKKQVLGVVVPNDLKKDLEE 659
QY 1128 AM-----LTLELERSALLQTVLELRSSAEPSPDREPECTOPEPTG 1167
Db 660 TMKKNTQVMDENKLDQRISELERLIEETANKNGTLEFYSNKNPLG 709

RESULT 13

T47641

hypothetical protein T15C9.20 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47641
R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24470
A:Accession: T47641
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-814 <MEW>
A:Cross-references: EMBL:AL132970
A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetics:
A:Map position: 3
A:Introns: 28/2; 51/3; 252/3; 312/3; 350/3; 370/2; 392/3; 410/1; 441/1; 469/1; 604/3;
A:Note: T15C9.20

Query Match 5.2%; Score 317; DB 2; Length 814;
Best Local Similarity 22.6%; Pred. No. 6.2e-10;
Matches 151; Conservative 108; Mismatches 220; Indels 188; Gaps 28;

QY 645 HLGDSGGTPVKTRR-HSWRQOIFLRVATPQKACDS--SSRYEDYSELGELPPSPLEPVCE 702
Db 26 HKRDAYGPPVRPQHVRQYREYADIVKEEERSDRWSSFLEDHVESTELPTNGSSSENI-- 83
QY 703 DGPFGPPPEKKRTSREL-----ELMQ 725
Db 84 HAPFS---ESEKEKEKELNKGPEDLHTDKGSDVTPDNASEEGHPDAEKVNRVQMT 140
QY 726 K-----AILQOILLRMEKENQKLOASENDLLNKLRLDYEEITPCLEKVTWVWKMST 780
Db 141 ETRPSLRSTEDLMSIRVKKGD-LSKSEQ-----APVK-ISPFDACKS 185
QY 781 PGRSRIK-----FDMK-----MHSVAGQGV 801
Db 186 KGASDIDSEDFYDVERSVDQGSSTSDGTGVSIGIPVAADASPLSTCPMKVEELVIRGV 245
QY 802 PRHHRGETWK-FLAQFHLKHQFPSPKQPKDVPYKELK-----QLTSOHA-----847
Db 246 PMALRGELWQAEVGVYR-----KRRKDY-YONLLAADSGSVNTIEQDMQHVDDKGS 295
QY 848 -----TLIDLGRTFPTHPYPSAQLGAGQLSYNLKAYSLDQEVGYCQGL 893
Db 296 STESAVVEKWKQIEKDLPRTFPGHPALDDD---GRNALRLLLTAYARHNSVGYCQAM 352

Qy 894 SFVAGILLHMSSEAEAFKMLFLMFDMLGRKQYRDPDMLILQIQMYQLSRLLDHYHRDLYN 953
Db 353 NFFAALLLLPEENAFWALIGLIDDY-FNGYISEEMIESQVDQLVLEELVRERPKLVH 411
Qy 954 HLEEEHIGPSLYAAPWFLTMFASQPLGFVARVFMIFLOQTEV-IFKVALSLLGSHKPL 1012
Db 412 HLDYLGQVAVWGPWFILSIFANMLPWSVLVWVDFEETRVMFLFTALAMELYGPA 471
Qy 1013 ILQHNLETVDFIK-----STLPNGLVOME-----KTINQVEMDI----- 1050
Db 472 LVTTKADGAVTLLQSLAGTSFSSQVLVLTACMGQYVHRIQLBELRSKHRPAVIAALEE 531
Qy 1051 -AKOLOAYEVE-----YHVLQEE-----LIDSSPLSDNQRWDKLEKTNSSLRKQNLDDLE 1099
Db 532 RSKGLQARWDSKGLASKYLNFKQDPKSVLVDKASLSNGSLSRSGSSNADEVLSVITG 591
Qy 1100 QIQVANGRIQSLAEATIE--KLTSSESKLQAMLTLELERSALLOTVESLRR--SAEPSD 1155
Db 592 DGEVDS--VDQLQAQAECLKLEEK--RSALLRAEELEIALMEIVKEDNRRQLSAKVEQ 646
Qy 1156 REPECTQ 1162
Db 647 LEQEMAE 653

RESULT 14

T16712

hypothetical protein R11B5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T16712

R:Leinbach, D.

submitted to the EMBL Data Library, December 1995

A:Description: The sequence of C. elegans cosmid R11B5.

A:Reference number: 218564

A:Accession: T16712

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-519 <LEI>

A:Cross-references: EMBL:U42831; NID:g1123122; PID:g1123123; PIDN:AAA83513.1; CESP:R11B5

A:Gene: CESP:R11B5.1

A:Introns: 13/2; 48/1; 85/2; 146/3; 190/2; 226/1; 248/3; 381/2; 473/3

Query Match 4.8%; Score 293; DB 2; Length 519;
Best Local Similarity 21.3%; Pred. No. 7.1e-09;
Matches 140; Conservative 103; Mismatches 237; Indels 178; Gaps 25;

Qy 497 KNAKRSLSLESILSRGNKARGLOEHSISVDLSDSSLSSTLSNTSKPSVCERKALPIS 556
Db 8 KNAAGLCLEPLCSL-----SLNTRADCSSSSESDQDQYDSGIGCTE-----S 49
Qy 557 ESSFKLLGSSDLSDESHPLEEPAPLSPOAFRRRANTLSHRPIECQPPQARGSPG 616
Db 50 DSKRSSLDRRSELSDFDP--PSKETLSKSRFF-----NFP----- 84
Qy 617 VSQRKLRYHVSSTETPIERKDFESKANHLGDSGTPVTKRRHSWRQIFLRVATPPQKAC 676
Db 85 -----KNFFSNKEKGKMGKSGNKGQVMATGL----- 116
Qy 677 DSSSRYEDYSELGELPPRSPLEPVCEGDFGPPPEKRTSRELWOKAILQOILLR 736
Db 117 -----ILEERPGLPSKSADEAAHH-----KOMYLDI 143
Qy 737 MEKENQKLOASNDLLN---KRLKLDYEITPC---LKEVTVWEKMLSTPGRSKIKEDM 790
Db 144 LEQAAKKQRAEKERLQAKAEQKRLSEQTAACHRVVWVEQILPKWDEMK-----DS 193
Qy 791 EKMHSVAGQVPRHRHGRWKF--LAEQFHLKHOPPSKQPKDVPYKELLKOLTSQOH--- 846
Db 194 KRCRELWQGVPAKVRGELWFLFTIGNQIEITKELYDGLM--DQAEKIAKQLAEQKNKSA 251

Qy 847 -----AILDLGRTPPT-----HPYFSAQLGAGLSLYNLKAYSLLDQEVGYCQ 891
Db 252 ERKETSVTQIIHLDAFRTFTSLGMFOKQDGPYD-----HLLKLSAYAILRPIGIVYQ 303
Qy 892 GLSFVAGILLHMSSEAEAFKMLKFL-MFDMGLRKQY-----RPMIILQIQMYQLSRLLDH 946
Db 304 SMTFAAVALLTQMDYPAF--ISFANLLDRSLQSAFFGLKQPMTEYFI---AYDRYLEQ 358
Qy 947 YHRDLYNLEHEHIGPSLYAAPWFLTMFASQPLGFVARVFMIFLOQTEVIFKVALSLL 1006
Db 359 ELPALHQHLDKLDVDPDLYLIEWTFAMYAKSLPDLVTCRIMDVYFRDGEELFKAAAGIL 418
Qy 1007 GSHKPLILQHENLETVDFIKSTLPN-LGLVQEMXTINQVEMDIKQALQAYEVEVHVLIQ 1065
Db 419 RMIEPKLITMD-FDCCVEFL-TKLPNTLTGAEFLFNIEPFMRPYNGESRSKRFSQIFQ 476
Qy 1066 BELIDSSPLSDNQRWDKLEKTNSSLRKQNLDDLEQLQVANGRIQSLAEATIEKLLSES 1123
Db 477 E-ID-----ERYNPGTGTARTQITHNVQELKMS-----KSLSGFIKDLLSSPS 518

RESULT 15

T48686

hypothetical protein DKFZp761D1823.1 - human

C:Species: Homo sapiens (man)

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000

C:Accession: T48686

R:Boecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, April 2000

A:Reference number: 224536

A:Accession: T48686

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-468 <AAA>

A:Cross-references: EMBL:AL353935

A:Experimental source: adult amygdala; clone DKFZp761D1823

C:Genetics:

A:Note: DKFZp761D1823.1

Query Match 4.8%; Score 290; DB 2; Length 468;
Best Local Similarity 24.9%; Pred. No. 8.9e-09;
Matches 121; Conservative 73; Mismatches 181; Indels 110; Gaps 17;

Qy 664 QIFLRVATPQKA-----CDSSRYEDYSELGELPPRSPLEPVCEGDFGPPPE 711
Db 20 KIWKRVAEKEKALLTKCAYLQARNQVSKY-----LAGRLRLQALGD----- 63
Qy 712 EKKRTSRELRELWOKAILQIILLRWKENQKLOASENDLLNKLKL-----DYE-EIT 764
Db 64 EASECSELLRLQVQEA-----LOWEAGEASSDSIELSPISKYDEYGLTVDPDYEDDL 116
Qy 765 PCLKEVTVWEKMLSTPGRSKIKFDMKMHSAVGQ-----GVPRHRHGEIWK 811
Db 117 KLLAKIQALESRSHLLGLEAVDRPLRERWALGDLVPSAELKQLLRAGVPREHRPVWR 176
Qy 812 FLAEQFHLKHOPPSKQPKDVP--YKELLKOLTSQOH-----AILDLGRTPFTPHYFSAQ 865
Db 177 WLW---HL-----RVQHLHTPGCYOELLSRGQAREHPAARQIOLDLNRTPFNKHTCP 227
Qy 866 LGAGLSLYNLKAYSLLDQEVGYCQGLSFVAGILLHMSSEAE-AFKMLKFLWDFMGLRK 924
Db 228 TSSFPDKLRRVLLAFSWQNPTIGYCOGLNRLAAIALLVLEESAFWCLVAIVETIMPAD 287
Qy 925 QYRPMIILQIQMYQLSRLLDHYHRDLYNLEHEHIGPSLYAAPWFLTMFASQPLGFVA 984
Db 288 YCNLTASQDQVRVLOQLLSEKLPRLMAHLGQHVDLSLVTFNFWFLVAFADSLISNILL 347
Qy 985 RVFDMIFLOQTEVIFKVALSLLGSHKPLILQHN---LETIVDFIKSTLPNGLVQMEKT 1041
Db 348 RWMDAFLVEGTVKVPVRYALATFKYNEKETLRLQNGLEIYQYLRFFTKTISN-----SRKL 402

QY 1042 INQVFENDIAKQLQAYEVEYHVLQEELIDSSPLSDNQ-----RMDKLEKTNNSSLRKQNLDL 1097
Db 403 MNTAFN-----DMNPFMKQLROLRMVHRERLEAELE----- 435

QY 1098 LEOLQ 1102
Db 436 LEOLK 440

Search completed: August 28, 2002, 16:24:25
Job time: 118 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:22:27 ; Search time 38.61 Seconds
(without alignments)
3360.119 Million cell updates/sec

Title: US-09-762-311-5

Perfect score: 6055

Sequence: 1 MEPIITFTARKHLLPNEVSD.....RSAEPSDRPECTQPEPTGD 1168

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

```

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6055	100.0	1168	AAV45096	Human TBC-1 protei
2	4976	82.2	1141	AAW44777	Human tbc-1 protei
3	4823	79.7	937	AAK38698	Human polypeptide
4	4772	78.8	981	AAW43195	Human ORFX ORF2959
5	3891	64.3	763	AAW40484	Human polypeptide
6	3463	57.2	674	AAW95675	Human protein sequ
7	1600.5	26.4	343	AAU21586	Novel human neopla
8	1559.5	25.8	390	AAU21586	Novel human neopla
9	1467.5	24.2	1379	ABW57823	Drosophila melanog
10	1334	22.0	265	AAW93840	Human polypeptide
11	588.5	9.7	1119	AAW40436	Human polypeptide

12	580	9.6	1069	22	AAW38650	Human polypeptide
13	580	9.6	1069	22	AAW38650	Human polypeptide
14	492	8.1	1194	22	ABW62985	Drosophila melanog
15	481	7.9	94	20	AAV11605	Human 5' EST secre
16	479.5	7.9	803	22	ABW65883	Drosophila melanog
17	392.5	6.5	353	21	AAW06209	Arabidopsis thalia
18	386	6.4	356	21	AAW06209	Arabidopsis thalia
19	382	6.3	292	21	AAW06210	Arabidopsis thalia
20	371.5	6.1	749	22	AAW39630	Human polypeptide
21	371.5	6.1	749	22	AAW39630	Human polypeptide
22	367.5	6.1	749	22	AAW78509	Human protein SEQ
23	367.5	6.1	749	22	AAW78509	Human protein SEQ
24	367.5	6.1	788	22	AAU17163	Novel signal trans
25	367.5	6.1	777	22	AAW78508	Human protein SEQ
26	367.5	6.1	796	22	AAW79492	Human protein SEQ
27	367.5	6.1	796	22	AAW79493	Human protein SEQ
28	367.5	6.1	796	22	AAW41415	Human polypeptide
29	367.5	6.1	796	22	AAW41416	Human polypeptide
30	366	6.0	298	21	AAW20060	Arabidopsis thalia
31	364.5	6.0	244	21	AAW06211	Arabidopsis thalia
32	347.5	5.7	907	22	AAW78535	Human protein SEQ
33	344	5.7	250	21	AAW20061	Arabidopsis thalia
34	343.5	5.7	893	22	AAW79519	Human protein SEQ
35	336	5.5	557	22	ABW07606	Novel human diagno
36	334	5.5	272	21	AAW86335	Human gene 2-encod
37	328.5	5.4	519	22	AAW40360	Human polypeptide
38	328.5	5.4	533	22	AAW95321	Human protein sequ
39	326	5.4	371	22	AAW42146	Human polypeptide
40	325.5	5.4	457	22	ABW63679	Drosophila melanog
41	325.5	5.4	571	22	ABW67288	Drosophila melanog
42	325	5.4	1092	22	ABW20243	Novel human diagno
43	319.5	5.3	533	22	ABW60453	Human cell cycle a
44	315.5	5.2	363	22	ABW61447	Drosophila melanog
45	315	5.2	508	22	AAW93726	Human polypeptide

ALIGNMENTS

RESULT 1

AAV45096

ID AAV45096 standard; Protein; 1168 AA.

AC AAV45096;

XX 31-MAY-2000 (first entry)

XX Human TBC-1 protein.

XX TBC-1; human; blallelic marker; chromosome 4; cell cycle regulator;

XX tissue differentiation; yeast regulator; BUB2; cdc16; tre2-oncogene;

XX linkage analysis; genetic map; detection; diagnosis; genotyping;

XX mitosis; prostate cancer; transgenic animal; screening; cytokinesis.

XX Homo sapiens.

XX Key

XX Domain

XX Location/Qualifiers

XX 786..974

XX /label= "TBC_domain

XX /note= "Regulates protein-protein interaction"

XX Region

XX 886..893

XX /note= "This sequence interacts with a kinase"

XX WO200008209-A2.

XX 17-FEB-2000.

XX 06-AUG-1999;

XX 99WO-IB01444.

XX 07-AUG-1998;

XX 98US-0095653.

XX (GEST) GENSET.

PI	Blumenfeld M, Bougueleret L, Chumakov I;				
XX	WPI: 2000-205736/18.				
DR	N-PSDB; AA250906, AA250907.				
XX	New isolated human TBC-1 nucleic acids, useful for developing products				
PT	for the diagnosis and treatment of disorders involving cell				
PT	proliferation, particularly prostate cancer				
XX					
XX	Claim 16; Page 162-165; 166pp; English.				
PS					
XX					
CC	The present amino acid sequence is the human TBC-1 protein, encoded by				
CC	exons 1-2 and A-L. TBC-1 protein is involved in the regulation of cell				
CC	cycle, mitosis, cytokinesis and tissue differentiation in mammals. The				
CC	TBC domain is homologous to regions in the tre2-oncogene and in the				
CC	yeast regulators of mitosis BUB2 and CDC16. TBC-1 gene is mapped to a				
CC	candidate region of prostate cancer on chromosome 4. An alteration of				
CC	TBC-1 sequence may be associated with a pathological condition, resulting				
CC	in abnormal cell proliferation leading to cancer, e.g. prostate cancer.				
CC	Biallelic markers present in the sequence can be used for generation of				
CC	genetic maps, linkage analysis and association studies. TBC-1 sequence				
CC	can be used for detection, diagnosis, genotyping, production of				
CC	transgenic animals and screening of compounds for use in therapy.				
XX					
SQ	Sequence 1168 AA:				
	Query Match 100.0%; Score 6055; DB 21; Length 1168;				
	Best Local Similarity 100.0%; Pred. No. 0;				
	Matches 1168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MEPTTARKHLLPNEVSDVDFGLQVLSLPVHSLTTPMLPWVAEVRRLSROSTRKEPV 60				
DB	1 MEPTTARKHLLPNEVSDVDFGLQVLSLPVHSLTTPMLPWVAEVRRLSROSTRKEPV 60				
QY	61 TKQVRLCVSPGLRCEPEGRSQWQDPLIYSSIFECKPQRVHKLIHNSHDPVSFACLIKE 120				
DB	61 TKQVRLCVSPGLRCEPEGRSQWQDPLIYSSIFECKPQRVHKLIHNSHDPVSFACLIKE 120				
QY	121 DAVHRQSIYVFVADDTQVPELISSIRQAGKIARQELHCPSEFDDTFSKKFEVLFQGR 180				
DB	121 DAVHRQSIYVFVADDTQVPELISSIRQAGKIARQELHCPSEFDDTFSKKFEVLFQGR 180				
QY	181 VTVAHKKAPPALIDECETENHNHVSGRGSESPRNPAAPTGQBPVRRPMKSFQSPQ 240				
DB	181 VTVAHKKAPPALIDECETENHNHVSGRGSESPRNPAAPTGQBPVRRPMKSFQSPQ 240				
QY	241 LRSIAFKELQDGLRSSGFFSFEESDIENHLISGHNIVQPTDIEENRTMLFTIQSEV 300				
DB	241 LRSIAFKELQDGLRSSGFFSFEESDIENHLISGHNIVQPTDIEENRTMLFTIQSEV 300				
QY	301 YLISPDTKKTALEKNFKEISFCOGIRHVDHFGFICRESSGGGFHFVCYVFOCTNEALV 360				
DB	301 YLISPDTKKTALEKNFKEISFCOGIRHVDHFGFICRESSGGGFHFVCYVFOCTNEALV 360				
QY	361 DETMNTLQAFVAAVQOTAKAPQAQCEGCPQLQSHKLCERIEGMNSSTKLELQKHLTT 420				
DB	361 DETMNTLQAFVAAVQOTAKAPQAQCEGCPQLQSHKLCERIEGMNSSTKLELQKHLTT 420				
QY	421 LTNOEQATIFEVOKLRNEQRENELIISFLRCLVEEKKQEHIIHIGEMKQTSQMAENI 480				
DB	421 LTNOEQATIFEVOKLRNEQRENELIISFLRCLVEEKKQEHIIHIGEMKQTSQMAENI 480				
QY	481 GSELPSPATFRDLMLKNKAKRSLTESLESILSRGNKARGLOEHSISVDLSDSLSTLSN 540				
DB	481 GSELPSPATFRDLMLKNKAKRSLTESLESILSRGNKARGLOEHSISVDLSDSLSTLSN 540				
QY	541 TSKEPSVCEKEALPISSESKLGSDDLSSDSESHLPEEPAPLSPQAPRRRANTLSHF 600				
DB	541 TSKEPSVCEKEALPISSESKLGSDDLSSDSESHLPEEPAPLSPQAPRRRANTLSHF 600				
QY	601 PIECQEPQPARSGVSGQRKLARYHSVSTETPHERKDFESKANHLGDSGGTVPVKTRRHS 660				

DB	601	piecqpqpqparsgvsgqrklmryhsvstetptherkdfeskanhlgdsggtpvktrrhs	660
QY	661	WRQOQFLRVATPQKACDSSRYEDYSELGELPPRSLEPVCEGDPGPPPEEKRTSREL	720
DB	661	WRQOQFLRVATPQKACDSSRYEDYSELGELPPRSLEPVCEGDPGPPPEEKRTSREL	720
QY	721	RELWQKAIQQLILLRMEKENQKQASENDLNKRLKLDYEETPCPKETVTWVKMLST	780
DB	721	RELWQKAIQQLILLRMEKENQKQASENDLNKRLKLDYEETPCPKETVTWVKMLST	780
QY	781	PGRSKIKFDMKMHSAVGVGVRPHHGETWKFIAEFGHLKHQPPSQQKQDVPYKELLKQ	840
DB	781	PGRSKIKFDMKMHSAVGVGVRPHHGETWKFIAEFGHLKHQPPSQQKQDVPYKELLKQ	840
QY	841	LTSQQAAILDILGRTPTHPYSAOLGAGOLSLYNILKAYSLDDQFVGVCOGLSFVAGIL	900
DB	841	LTSQQAAILDILGRTPTHPYSAOLGAGOLSLYNILKAYSLDDQFVGVCOGLSFVAGIL	900
QY	901	LHMSDEEAFKMLKFLMFDMGKRGKQYRPMIIFIQMYQLSRLLDHYHRDLYNHLSEHEI	960
DB	901	LHMSDEEAFKMLKFLMFDMGKRGKQYRPMIIFIQMYQLSRLLDHYHRDLYNHLSEHEI	960
QY	961	GPSLYAAPFLTNFASQFPLGFVAVDFMIFLOGTEVIFKVALSLGSHKPLILQHENLE	1020
DB	961	GPSLYAAPFLTNFASQFPLGFVAVDFMIFLOGTEVIFKVALSLGSHKPLILQHENLE	1020
QY	1021	TIVDFIKSTLPLNLGVOMEKTIQVEMDIQAKOLQAEYEVHVLQBELIDSSPLSDNORM	1080
DB	1021	TIVDFIKSTLPLNLGVOMEKTIQVEMDIQAKOLQAEYEVHVLQBELIDSSPLSDNORM	1080
QY	1081	DKLEKTNSSIRKQNLDLLEOLOQVANGRIOSLEATIEKLSSSESKLKQAMLTLELERSALL	1140
DB	1081	DKLEKTNSSIRKQNLDLLEOLOQVANGRIOSLEATIEKLSSSESKLKQAMLTLELERSALL	1140
QY	1141	QTVVEELRRSAEPSDREPECTQPEPTGD 1168	
DB	1141	QTVVEELRRSAEPSDREPECTQPEPTGD 1168	
RESULT	2		
AAW44777			
ID	AAW44777	standard; Protein; 1141 AA.	
XX	AAW44777;		
AC	AAW44777;		
DT	01-JUN-1998	(first entry)	
XX	Human Tbc-1 protein.		
DE	tre-2; BUB2; cdc16; mast cell; probe; hybridisation; subtraction method;		
KW	mouse; transcription factor; differentiation; proliferation; human;		
KW	acute myelogenous leukaemia.		
XX	Homo sapiens.		
OS	US5700927-A.		
XX	23-DEC-1997.		
PD	23-DEC-1994; 94US-0363300.		
XX	23-DEC-1994; 94US-0363300.		
PF	(CHIL-) CHILDRENS MEDICAL CENT.		
XX	Richardson P, Zon L;		
PI	WPI; 1998-062437/06.		
DR	N-PSDB; AAV05886.		
XX	DNA encoding Tbc1 polypeptide - useful for treating leukaemia		
PT	Claim 1; Fig 1A-B; 22pp; English.		
PS			

XX This is the amino acid sequence of a Tbc1 (tre-2, BUB2 and cdc16)
 CC polypeptide, isolated from a bone marrow-derived mast cell DNA library.
 CC The screen was carried out using a probe generated by a subtraction
 CC method which compared mRNA expression in an undifferentiated mast cell
 CC line P815 and cell line PC76 (P815 cells transfected to express the
 CC murine GATA-1 transcription factor - a factor which controls the
 CC expression of genes involved in mast cell differentiation). Tbc1 encodes
 CC a protein involved in the coupling of cell proliferation to cell
 CC differentiation, which can be used to treat leukaemia (especially acute
 CC myelogenous leukaemia) by causing leukaemic cells to differentiate.
 XX
 SQ Sequence 1141 AA;

Query Match 82.2%; Score 4976; DB 19; Length 1141;
 Best Local Similarity 87.9%; Pred. No. 0;
 Matches 970; Conservative 47; Mismatches 74; Indels 12; Gaps 5;

QY 37 MPMLPVVAEVRRLSRQSTRKEPTVKQVRLCVSPSLRCEPEGRSQOWDPLIYSIFEC 96
 Db 1 mpmlpvvaevrrlsgqskkprtkqvrilwvsgslrcepdiekspwdplicsifec 60
 QY 97 KPOKVKHLIHNHSDPSYFACILKEDAVHROSICVVFKADDTQKVPETIISIROAGTIARQ 156
 Db 61 kprvkhlihnshdpsyfacilkedaaahrsqicyvfakddqtkvpelissirgagkiarq 120
 QY 157 EELHCPSEFDFTSKKFEVLFCGRVTVVAHKKAPPALIDECIERFNHVSGRSGESPRNP 216
 Db 121 eelrcpsefdftfakkfevlfcgrvrvvahnkappalideciekfhnvscgrtd----- 174
 QY 217 PHAAPTGS-QEPVVRPMRKSFSGPLASLAFRKLQDGLRSGGFFSPESDIENHLIS 275
 Db 175 -weaptgqseaprpnrksfsgplrsrlafkqfcdaslrss-tfssf-dndienhlig 231
 QY 276 GHNVQPTDEENRTMLFTIGQSEVYLISPDTKKIALEKFKETSCQSGIRVWDHFGFI 335
 Db 232 ghnvqptdneenrtmlftigpsevylispdtkkialekfnkelsfcsqgirvhdhfgfi 291
 QY 336 CRESSGG--GGFHFVCVFOCTNEALVDEIMMTLKQFTVAAVQOTAKAPAQICEGGLQ 393
 Db 292 cressggsgghfvcyvfqctnealvdeimmtlkqftvaavqgtakapaqicegglq 351
 QY 394 SLHLKCEIRIEGMNSKTKLEQLHLLTTLTNOEQATIFEVQKLRPRNEORENELISFLR 453
 Db 352 glhlcieriegmnskkleqlhlltlttngqeatifeevqkrlprnrenerelislflr 411
 QY 454 CLYEKQKEHIHIGEMKQTSQMAENIGSELPPSATFRDLMLKNKAKRSLTESLESILS 513
 Db 412 clyeekqeshgtgpkqtlqvaenlgsdlppsasfrldsknrakrslteslesils 471
 QY 514 RGNKARGLOEHSISVLDLSSLSLTSNTSPKSPVCEKEALPISESPKLLGSSSEDLSSDS 573
 Db 472 rgnkarglqdhssavldlssstsltsntskelsmgkkaefpvssetsfkllgssddlsds 531
 QY 574 ESHLPEEPAPLSPOAFRRRANTLSHPPIECQEPPOARGSPGVOSKLMRYHSVSTETP 633
 Db 532 eghlaeesallspqafrrrantlshpvecpappespgvsgvrklmryhsvstetp 591
 QY 634 HERKDFESKANHLGDSGGTPVKVRRHSWROQIFLRVATPQKADSSSRDYEDYSELGLPP 693
 Db 592 herkdfeskanhlgdtdgtpvkttrhsvrqqlfrvatpdkadsparyedyseelgealp 651
 QY 694 RSPLEPVCEDPGFPPEPKKRTSRELRLELWQKAILQOILLRMEKENQKLAASENDLLN 753
 Db 652 rsplepvcedgpgfpykkrgrthasfescgkprscsrclvrmekenqklqasendlln 711
 QY 754 KRLKLDYEETPCLEKVTTVWERMSTPGRSKIKFDMKMHSAVQGVPPRHHRGEITWFL 813
 Db 712 krlkldyeetpclekvttvwermstpggrskikfdmekvhsavggvpprhrgeitwfl 771
 QY 814 AEQFLKHKQPPSKOOPKDPVKELKLTQSOHAILDLGTRTPPTPHYFSAQLGAGLQSL 873

Db 772 aeqfhkhkpfspkqpkdvpvykellkkltsqgqailldgrtftphpyfsaqigagqlsl 831
 QY 874 YNLKAYSLLDQVGVYCGQLSFSVAGILLHMSBEEAFKMLKFLMDFMGLRQYRPDMIIIL 933
 Db 832 ynlkaysllldqvgycqqlsfvagillhmseeafkmlkflmfdnglkrqyrpdmiiil 891
 QY 934 QIQMYQLSRLLDHYHRLDLYNHLHEEIGPSLYAAPWFLTMFASQPLGTFVARVDFMIFLQ 993
 Db 892 qiqmyqlsrlldhyrdlynhleehetgptyaapwfltfvasqplgtfvarvdfmiflq 951
 QY 994 GTVIFKVALSLGSHKPLILOHENLETIVDFIKSTLPNLGLVQMEKTIINQVPEMDIAQ 1053
 Db 952 gsevfkaivslslgshkplilqhenletivdfikntlpnlglvqmektisqvfemdiakg 1011
 QY 1054 LQAYEYHVHLQBELIDSSPLSDNQRMKLEKTNSSLRKONLLEQLQVANGRIQSLA 1113
 Db 1012 lqayeveyhvvgqeellssplsdnqrmelektntslrkqldleqlqvanarigalea 1071
 QY 1114 TIEKLSSSESKLQAMLTLELER 1136
 Db 1072 tveklitssesklqkqaltiever 1094

RESULT 3
 AAM38698
 ID AAM38698 standard; Protein; 937 AA.
 XX
 XX AAM38698;
 XX AC
 XX 22-OCT-2001 (first entry)
 XX DT
 XX Human polypeptide SEQ ID NO 1843.
 DE
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 XX Homo sapiens.
 OS
 XX WO200153312-A1.
 XX PN
 XX 26-JUL-2001.
 XX PD
 XX 26-DEC-2000; 2000WO-US34263.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSB-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AAI57854.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 3; SEQ ID NO 1843; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ

Sequence 937 AA;

Query Match 79.7%; Score 4823; DB 22; Length 937;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 937; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 MRKFSQGLRLAFKELQDGLRSGFFSFESDIENHLISGHNIYQPTDIEENRTM 291
Db 1 mrksfqpglrsafakelqdgglrsgffsfesdienhlisghniyqptdienrtm 60
QY 292 LFTIGSEVYLISPDFFKTALENKFEISFCSGIHRVDFHFGICRESSGGGFHVYCV 351
Db 61 lftigsevyllspdkkaleknfkseicqgrrhvdhfgicressgggghfvycv 120
QY 352 FQCTNEALVDEIMMTLKQFTVAQVQOTAKAPQALCEGCPLOSLHLKCRIEGMNSKTK 411
Db 121 fctnealvdelmmtlkqftvaavqqtakapqalcegcplqslhlcleriegmskkt 180
QY 412 LEQKHLTLTNDQEQATIEEVOKLRPRNEQRENEILISFLRCLYBEKKEHIIHGEMQ 471
Db 181 lelqkhltlndeqatifeevqklprneqreneilislrclyeekqehihgemq 240
QY 472 TSQMAAENIGSELPPSATFRDMLKNKAKRSITLESILSRGNKARGLOEHSISVDLD 531
Db 241 tsqmaeenigselppsatfrdmlknkaksitesilsrgnkargloehsisvld 300
QY 532 SLSLSTLSNTSKRPSVCEKEALPISSESPKLLGSSDLSSESHPLEPAPLSPQOAFR 591
Db 301 slslstlntskpsvcekealpisessfkllgssedlsdeshlpeepaplsqqafr 360
QY 592 RRANTLUSHPICEQPPPARSGVQSKLMRYHSVSTETPHERKDFESKANHLGDSGG 651
Db 361 rrantlshfpieqppparsgvsgqrklmryhsvtetptherkdfeskanhlgdsg 420
QY 652 TPVKTRRHNRQOIFLRVATPOKACDSSRYEDYSELGELPPRSPLEPCEDGPGPPPE 711
Db 421 tpvktrrhnrqoiflrvatpqacdsrriedysegelpprsplepcedgpgpppe 480
QY 712 EKKRTSRELWQKAILQOILLRMKENQKLAQSENDLLNKKRLDYEEITPCLKEVT 771
Db 481 ekkrtsrelwqkailqoillrmkenqklqasendllnkrkldyeetpcлкеvt 540
QY 772 TWVEKMLSTPGRSKIFDMEKMSAVGQVPRHHRGEIWKFLAEQFHLKHQPPSKQPKD 831
Db 541 twvekmlstpgsrskifdmekmhsavggvprhrhrgewkflaeqfhlkhqpskqpkd 600
QY 832 VPYKELLKOLTSQOAILDLGRTFTTPVFSQALGAGOLSLYNILKAYSLLDQEVGYCQ 891
Db 601 vpykellkoltseqhaildlgrtfttpvfyseaqagqlslynilkayslldqevgycq 660
QY 892 GLSFVAGILLHMSBEEAFKMLKFLMDFMDGLRKQYRPDMIIQIQYQISRLLDHYRDL 951
Db 661 glsfvagillhmsbeefkmlkflmfdmglrkqy:pdmiliqyqisrllldhyrdl 720
QY 952 YNLEHEHGIPSLIYAPWFLTWPAOFPGLCFVARVDMIFLQCTEIVFKVALSLGSHKP 1011
Db 721 ynleehhgipsliyaapwfltwpaofpglcfvarvdmiflqgteivfkvalsllgshkp 780

QY 1012 LILQHENLETIYDFIKSTLPNLGLVQMEKTIQVFPENDIAKQLQAYEVHYHVLQELIDS 1071
Db 781 lilqhenletivdfikstlpnlglvqmektiqvfemdiakqlqayevhyhvlqeelids 840
QY 1072 SPLSDNORMDKLEKNSSLRKQNLDLLEQLQVANGRIQSLEATIEKLSSSESKLKQAMLT 1131
Db 841 splsdnqrmdkлектnsslrkqndlleqlqvangrigsleatieklssesklqamlt 900
QY 1132 LELESAALQTVVEELRRRSAPESDREPECTQPEPTGD 1168
Db 901 lelersallqtveelrrrsaeptsdrepectqpeptgd 937

RESULT 4

AAB43195
ID AAB43195 standard; Protein; 981 AA.

XX AAB43195;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2959 polypeptide sequence SEQ ID NO:5918.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO2000058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC77404.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 5097-5099; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAC40237 to AAC43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 763 AA;

Query Match 64.3% Score 3891; DB 22; Length 763;
 Best Local Similarity 99.7% Pred. No. 0;
 Matches 756; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 409 KTKLEQLHLLTNOQATFEVQKLRPRNEORENELIISFLRCLYEKQKEHIIHIGE 468
 DB 4 ktkleqlhlltlnqeqatfeevqklrprneoreneliisflrcifeekqkehnhige 63

QY 469 MKQTSQMAENIGSELPSPATRFRLMLKNKAKRSLTESLESILSRGNKARGLOEHSISV 528
 DB 64 mkqtsqmaenigseelpspatrfldmlknkaksrleslesilsrngnkarglqehsiv 123

QY 529 DLDSLSSTLSNTKEPSVCEKEALPISSESKLLGSSEDLSSDSHLPPEEPAPLSPQ 588
 DB 124 dlsslsstlsntkepsvcekealpisesskllgsseidlssdseshlpeepaplsppq 183

QY 589 AFRRRANTLSHPIECOEPPQARGSPGVQKLMRYHSVSTETPHERKDFESKANHLGD 648
 DB 184 afrrrantlshpiecpeppargspgvqklmryhsvstetpherkdfeskanhlgd 243

QY 649 SGGTPVTRHRSWQOJFLRVATPQKACDSSSRVEDYSELGELPPRSPLEPVCEDGFGP 708
 DB 244 sggtpvtrhswrqjflrvatpqkacdssryedyselgelpprsplepvcgedgfgp 303

QY 709 PPEKKTSTRELRLWQKAILQIILLRMKEKNOKLOASENDLNLKRLDYEBITPCLK 768
 DB 304 ppeekktstrelrlwqkailqillrmekendklasesndlnlkrldyeeitpclk 363

QY 769 EVTVWPKMLSTPGRSKIKEDMEKMHSAVGQVPRHRRHGEIWKFLADQFHLKHQFPSSKQ 828
 DB 364 evttwepkmlstpgsrskikfedmekmhavgvprhrrhgeiwlkfladqfhlkhqfsskq 423

QY 829 PKQVPYKELLKQTSQOHAIIIDLGRFPFPPYFSAOLGAGOLSLYLKAYSLLDQEVG 888
 DB 424 pkdvpkellkqtsqghailldgrtftphysaqlgagqlslylilkayslldqevg 483

QY 889 YCQGLSVAGILLHMESEAFKMLKFLMDFMGLRKQYRPMIILQITQMLSLLDHYH 948
 DB 484 ycqglsvagillhmseefkmlkflmfdmglrkqyrdpmiilqitqmlslldhyh 543

QY 949 RDLYNHLEEHIGPSLYAAPWFLTMFASQPLGFVARVDMIFLQGTVEVIFKVALSLGS 1008
 DB 544 rdlynhleehigpslyaaawfltmfasqplgfvavrdmiflqgtvevifkvalslgs 603

QY 1009 HKPILLOHENLETIVDFIKTLPNLGLVOMEKTIQVFMENDIAKQLOAYEYHVHLOEEL 1068
 DB 604 hkpilohenletivdfiktlpnlglvomekti qvfmendia kqlqayeyvhvloeel 663

QY 1069 IDSSPLSDNORMDKLENTSNLRKQNLDLLEQLQVANGRIQSLEATIEKLSSSKLKQA 1128
 DB 664 idssplsdnordmklentsnlrkqnlldleqlqvangriqsleatieklssesklkqa 723

QY 1129 MLTLELERSALLQVVEELRRRSAPSDREPECTGPTGD 1168
 DB 724 mltlelersallqveelrrrsakpsdrepectgptgd 763

RESULT 6
 AAB95675
 ID AAB95675 standard; Protein; 674 AA.
 XX
 AC AAB95675;
 XX
 DT 26-JUN-2001 (first entry)
 XX

DE Human protein sequence SEQ ID NO:18466.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 18466; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 674 AA;

Query Match 57.2% Score 3463; DB 22; Length 674;
 Best Local Similarity 100.0% Pred. No. 5.7e-269;
 Matches 674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 MLKNKAKRSLTESLESILSRGNKARGLOEHSISVLDLSSLSSTLSNTSKPSVCEKEALP 554
 DB 1 mlknkaksrleslesilsrngnkarglqehsivldlsslsstlsntskpsvcekealp 60

QY 555 ISSESSFKLLGSSEDLSSDSHLPPEEPAPLSPQAFRRRANTLSHPIECOEPPQARGS 614
 DB 61 isessfkllgsseidlssdseshlpeepaplsppqafrrrantlshpiecpeppargs 120

QY 615 PGVSQRKLMRYHSVSTETPHERKDFESKANHLGDSGGTGPVKTRHRSWQOJFLRVATPQK 674
 DB 121 pgvsqrklmryhsvstetpherkdfeskanhl gds ggtgpvktrhswrqjflrvatpqk 180

Qy	675	ACDSSRYEDYSELGELPPRSPLEPCVBCDGPFGPPPEKKRTSRELRLWOKAILQOILL	734	PR	07-JUL-2000;	2000US-0216647.
				PR	07-JUL-2000;	2000US-0216880.
Db	181	acdsrsyedyselgelpprplpvcvcdgpgpppekkrtrelrelwqkallqill	240	PR	11-JUL-2000;	2000US-0217487.
				PR	11-JUL-2000;	2000US-0217496.
Qy	735	LRMEKENQKLOASENDLNLKRLKLDYBEITPCLEKVTTVWEKMLSTPGRSKIKFDMKMH	794	PR	14-JUL-2000;	2000US-0218290.
				PR	26-JUL-2000;	2000US-0220963.
Db	241	lrmekeqkqasendlnlrlkldyeeitpclkevttvwekmlstpgskikfdmekmh	300	PR	26-JUL-2000;	2000US-0220964.
				PR	14-AUG-2000;	2000US-0224518.
Qy	795	SAVGQVPRHRHGRGIWFLAQFHLKHQFPKQOPKDPVPYKELLKQLTSCQHAILIDLGR	854	PR	14-AUG-2000;	2000US-0224519.
				PR	14-AUG-2000;	2000US-0225213.
Db	301	savgqgvprhrhgrgelwflaeqfhlkhqfksqpkdpvykellkqltsqghailidgr	360	PR	14-AUG-2000;	2000US-0225214.
				PR	14-AUG-2000;	2000US-0225266.
Qy	855	TFPTHYFSAQAGQSLYNILKAYSLLDQEVGYCGSLFVAGILLHMESEEAFAKMLK	914	PR	14-AUG-2000;	2000US-0225267.
				PR	14-AUG-2000;	2000US-0225268.
Db	361	tfpthyfsaqagqslsynilkayslldqevgycgslfvagillhmseeeafkmlk	420	PR	14-AUG-2000;	2000US-0225270.
				PR	14-AUG-2000;	2000US-0225477.
Qy	915	FLMFDMLRKQYRPMILQITQMTQLSRLLDYHRLDLYNHLHEHIGPSLYAAPFLTMF	974	PR	14-AUG-2000;	2000US-0225757.
				PR	14-AUG-2000;	2000US-0225758.
Db	421	flmfdmglrkqyrdpmilqltqmtqlsrlldyhrldlynhleehigpslyaaapfltmf	480	PR	14-AUG-2000;	2000US-0225759.
				PR	18-AUG-2000;	2000US-0226279.
Qy	975	ASQPLGVARVDFMIFLQGFTEVIFKVALSLGSHKPLILQHENLETIVDFIKSTLPNLG	1034	PR	22-AUG-2000;	2000US-0226681.
				PR	22-AUG-2000;	2000US-0226681.
Db	481	asqplgvarvdfmiflqgftevfikvalsllgshkplllqhenletivdfikstlpnlg	540	PR	22-AUG-2000;	2000US-0227182.
				PR	23-AUG-2000;	2000US-0227009.
Qy	1035	LVOMEKTIQVFMEDIAKQLOAYEVEYHVLQOEELIDSSPLSDNQRMDKLEKTNSSLRKQN	1094	PR	30-AUG-2000;	2000US-0228924.
				PR	01-SEP-2000;	2000US-0229287.
Db	541	lvomektiinqvfemediakqlqayevyhlqeelidssplsdnqrmdklektnsslrkqn	600	PR	01-SEP-2000;	2000US-0229343.
				PR	01-SEP-2000;	2000US-0229344.
Qy	1095	LDLLEQLOVANGRIOSLEATEIKLSSSKLKQAMLELERSALLQTVELRRRSAPPS	1154	PR	01-SEP-2000;	2000US-0229345.
				PR	05-SEP-2000;	2000US-0229509.
Db	601	ldlleqlqvangriqslsleatiekllssesklkqamlelersallgtveelrrrsaepp	660	PR	05-SEP-2000;	2000US-0229513.
				PR	06-SEP-2000;	2000US-0230437.
Qy	1155	DREPECTQPEPTGD 1168		PR	06-SEP-2000;	2000US-0230438.
				PR	08-SEP-2000;	2000US-0231242.
Db	661	drepectqpeptgd 674		PR	08-SEP-2000;	2000US-0231243.
				PR	08-SEP-2000;	2000US-0231244.
				PR	08-SEP-2000;	2000US-0231413.
				PR	08-SEP-2000;	2000US-0231414.
				PR	08-SEP-2000;	2000US-0232080.
				PR	08-SEP-2000;	2000US-0232081.
				PR	12-SEP-2000;	2000US-0231968.
				PR	14-SEP-2000;	2000US-0232397.
				PR	14-SEP-2000;	2000US-0232398.
				PR	14-SEP-2000;	2000US-0232399.
				PR	14-SEP-2000;	2000US-0232401.
				PR	14-SEP-2000;	2000US-0232401.
				PR	14-SEP-2000;	2000US-0233063.
				PR	14-SEP-2000;	2000US-0233064.
				PR	14-SEP-2000;	2000US-0233065.
				PR	21-SEP-2000;	2000US-0234223.
				PR	21-SEP-2000;	2000US-0234274.
				PR	25-SEP-2000;	2000US-0234997.
				PR	25-SEP-2000;	2000US-0234998.
				PR	26-SEP-2000;	2000US-0235484.
				PR	27-SEP-2000;	2000US-0235834.
				PR	27-SEP-2000;	2000US-0235836.
				PR	29-SEP-2000;	2000US-0236327.
				PR	29-SEP-2000;	2000US-0236367.
				PR	29-SEP-2000;	2000US-0236368.
				PR	29-SEP-2000;	2000US-0236369.
				PR	29-SEP-2000;	2000US-0236370.
				PR	02-OCT-2000;	2000US-0236802.
				PR	02-OCT-2000;	2000US-0237037.
				PR	02-OCT-2000;	2000US-0237038.
				PR	02-OCT-2000;	2000US-0237039.
				PR	02-OCT-2000;	2000US-0237040.
				PR	13-OCT-2000;	2000US-0239935.
				PR	13-OCT-2000;	2000US-0239937.
				PR	20-OCT-2000;	2000US-0240960.
				PR	20-OCT-2000;	2000US-0241221.
				PR	20-OCT-2000;	2000US-0241785.
				PR	20-OCT-2000;	2000US-0241786.
				PR	20-OCT-2000;	2000US-0241787.

RESULT 7

AAU21588

ID AAU21588 standard; Protein: 343 AA.

XX AAU21588;

XX 04-DEC-2001 (first entry)

XX DE

XX Novel human neoplastic disease associated polypeptide #21.

XX Human; neoplastic disease associated polypeptide; cancer;

XX hyperproliferative disorder; neural disorder; immune system disorder;

XX muscular disorder; reproductive disorder; gastrointestinal disorder;

XX pulmonary disorder; cardiovascular disorder; renal disorder;

XX neuroprotective; cytostatic; anti inflammatory; vasotropic.

XX Homo sapiens.

XX WO200155163-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01358.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

20-OCT-2000; 2000US-0241808.
 20-OCT-2000; 2000US-0241809.
 20-OCT-2000; 2000US-0241826.
 01-NOV-2000; 2000US-0244617.
 08-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246478.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246609.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249264.
 17-NOV-2000; 2000US-0249265.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 01-DEC-2000; 2000US-0250160.
 01-DEC-2000; 2000US-0250391.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI: 2001-465558/50.
 N-PSDB; AAS34787.
 Novel polypeptides and polynucleotides useful as diagnostic reagents to
 diagnose diseases or disorders associated with aberrant expression or
 activity of polypeptides, and for treating cancers, rheumatoid
 arthritis -
 Claim 11; SEQ ID No 315; 687pp; English.
 The present invention relates to the isolation of novel human neoplastic
 disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
 sequences encoding for these polypeptides. The sequences of the
 invention are useful in the diagnosis, treatment, prevention and/or
 prognosis of disorders involving neoplastic disease such as
 hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder

CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
 CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
 CC also be useful for treating other disorders such as neural disorders,
 CC immune system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC and renal disorders. The polynucleotide sequences of the invention are
 CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
 CC neoplastic disease associated polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 343 AA;
 Query Match 26.4%; Score 1600.5; DB 22; Length 343;
 Best Local Similarity 96.4%; Pred. No. 7.1e-120;
 Matches 320; Conservative 2; Mismatches 5; Indels 5; Gaps 1;
 QY 837 LKQLTSQOHAILDIDGRTPPTHPYSAOLGAGQLSLYNILKAYSLLDQEVGYCQGLSFV 896
 DB 17 lllql-----icilftgrtftphysaqlgagqlslnilkayslldqevgycqlsfv 71
 QY 897 AGILLHMSEEEAFKMLKFLMFDGMRKQYRPMIILQOMYOLSLRLLHDYHRDLNHL 956
 DB 72 agillhmseeeafkmlkflmfdmglrkgyrpdmiilqmqylslrllhdyhrdlynhle 131
 QY 957 EHEIGPSLYAAPFLTMEASOPFLGFVARVDFMIFLQGTVEIFKVALSLGSHKPLILOH 1016
 DB 132 eheigpslyAAPFLTMEASOPFLGFVARVDFMIFLQGTVEIFKVALSLGSHKPLILOH 191
 QY 1017 ENLETIVDFIKSTPLNLGLVQMEKTIQVFMEDIAKQAYEYHVQLBELIDSSPLSD 1076
 DB 192 enletivdfikstplnlglvqmektiqvfemdiaqkqayevyhlqelidssplsd 251
 QY 1077 NORMDKLEKTNSSLRKONLDLLEQLQVANGRIOSLEATIEKLSSSKLKQAMLTLELER 1136
 DB 252 ngrmdklektnsslrkgnldlleqlqvangrigsleatieklssesklqamltieler 311
 QY 1137 SALLQTVELRRRSAPSDREPECTOPEPTGD 1168
 DB 312 sallqtveelrrrsapdsrepectptgd 343
 RESULT 8
 AAU21586
 ID AAU21586 standard; Protein; 390 AA.
 XX
 AC AAU21586;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Novel human neoplastic disease associated polypeptide #19.
 XX
 KW Human; neoplastic disease associated polypeptide; cancer;
 KW hyperproliferative disorder; neural disorder; immune system disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW neuroprotective; cytostatic; anti inflammatory; vasotropic.
 XX
 OS Homo sapiens.
 XX
 PN WO200155163-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01358.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-02559678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465558/50.

N-PSDB; AAS34785.

Novel polypeptides and polynucleotides useful as diagnostic reagents to
 diagnose diseases or disorders associated with aberrant expression or
 activity of polypeptides, and for treating cancers, rheumatoid
 arthritis

Claim 11; SEQ ID No 313; 687pp; English.


```
Oy 558 SSFKLL--GSSEDLSSSESHLPREPAPLSPOQAQRRRANTLSHPTTECQEPPOPARGSP 615
Db 554 hnldiregsaeplgtq-----spgefrrsntvtgasp-----sskp 591
Oy 616 GVSQRK--LMRYHSVSTPHERKDFESKANHLGDSGTPVKTTRHSWROQIFLVRATPQ 673
Db 592 taeklkspmdiflkvgnspk-----aethgswrqailnsvtpps 633
Oy 674 KACSSSRDYEDYSELGLPPPSLEPVCEDPGFPPEKRTSRELRLWOKAILQOIL 733
Db 634 kglidsevtetfls-----pmckpakrgkrdaaelwrtairqtim 675
Oy 734 LLRWEKENQKLOASENDLLNKRKLKLDVEETPCLEKVEYTWVEKML---STP-GRSKTKFD 789
Db 676 lnrmtenamllqarqnenelrkldlyeeivpcdkqlerweqiernstqignnk---d 732
Oy 790 MEKMSAVGQGVPRHHRGEIKWFLAEQPHLKHQFP--SKQOPK-DVPYKELLKOLTSQOH 846
Db 733 pkvighairtyprskrgdvvtflaeq-hsmmtapvdkrfpnfntpyhmlklhtebqh 791
Oy 847 AILDLGRFTFTHFYFSAQLAGOLSLYNLIKAYSLLDQEVGYCOGLSFVAGIILLHMS 906
Db 792 aifidlgtrfphgkydplglglqlsfnllkayslldpelgycgglgfcgvllhcd 851
Oy 907 EEAFFKMLKFLMDGLKQOYRPMIILQIQMYQLSRLLHRYRDLNHLHEEIGPSLYA 966
Db 852 ansqglkhlmfrnmrkylpdmnkkqglqylsrivkxhdpdyvldqndvsptiya 911
Oy 967 APWELTFASQFPLGFVARVDFMIFLQGTVEIFKVALSLLGSHKPLILQHENLETIVDFI 1026
Db 912 apwiltvfssqfplgfvarvdfllflessvifkfaiallsvhkqllakdnfeelm 971
Oy 1027 KSTLPNLGLVOMETINQVFEKDIKQLOAYEVEYHVLQELIDSSPLSDNQRMDKL--E 1084
Db 972 ktvvpkmehtcmeglmklvfsmldgkqlaeYkveynvlqesi-----ttnhhlemnre 1026
Oy 1085 KTNSSLRKQNDLLEQLQVANGRIQSLEATIEKLLSSKQKQAMLTLELERSALLQTV 1144
Db 1027 kt-----qnghleqqkfagdsaqlett-----rssqqaqtlqsgvqseletiq 1073
Oy 1145 ELRRRSAPSDREPECTQP 1163
Db 1074 tlgyvgglvehnpdelp 1092

RESULT 10
AA093840
ID AA093840 standard; Protein; 265 AA.
AC AA093840;
XX
XX
XX
XX 06-NOV-2001 (first entry)
XX Human polypeptide, SEQ ID NO: 3915.
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
XX EP1130094-A2.
XX
XX
XX 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
XX 11-JAN-2000; 2000JP-0118774.
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T. Nishikawa T. Isogai T. Hayashi K. Ishii S. Kawai Y.
XX Wakamatsu A. Sugiyama T. Nagai K. Kojima S. Otsuki T. Koga H.
```

```
XX
DR WPI; 2001-524255/58.
DR N-PSDB; AAK94796.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 3915; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 265 AA;

Query Match 22.0%; Score 1334; DB 22; Length 265;
Best Local Similarity 100.0%; Pred. NO. 1.2e-98;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 904 MSEEAFKMLKFLMDGLKQOYRPMIILQIQMYQLSRLLHRYRDLNHLHEEIGPS 963
Db 1 mseeafkmlkflmdglrkqyrdmliiqmyqlsrllhryrdlnhlheehgpps 60

Oy 964 LYAAPWELTFASQFPLGFVARVDFMIFLQGTVEIFKVALSLLGSHKPLILQHENLETIV 1023
Db 61 lyaaapwiltvfassqfplgfvarvdfmiflqgtveifkvalsllgshkplliqhenletiv 120

Oy 1024 DFIRSTLPNLGLVOMETINQVFEKDIKQLOAYEVEYHVLQELIDSSPLSDNQRMDKL 1083
Db 121 dfirstlpnlglvomektinqvfekdiakqlayevyvhvlqeeldssplsdnqrmdkl 180

Oy 1084 EKTNSLRKQNDLLEQLQVANGRIQSLEATIEKLLSSKQKQAMLTLELERSALLQTV 1143
Db 181 ektnsslrkqndlleqlqvangrigsleatiekllssesklkqamltlelersallqtv 240

Oy 1144 EELRRRSAPSDREPECTQP 1168
Db 241 eelrrrsapdsrepectqptgd 265

RESULT 11
AA040436
ID AA040436 standard; Protein; 1118 AA.
XX
XX AA040436;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 5367.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX WC200153312-A1.
XX
XX 26-JUL-2001.
XX
```

PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 23-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AA159592.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Example 2; SEQ ID NO 5367; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA138642-AA142213) with neurotropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities, chemotactic/chemokinetic activity, haemostatic
 CC activin/inhibin activity, cancer diagnosis and therapy, drug screening,
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1118 AA;

Query Match 9.7%; Score 588.5; DB 22; Length 1118;
 Best Local Similarity 24.5%; Pred. No. 8.2e-38;
 Matches 229; Conservative 137; Mismatches 311; Indels 257; Gaps 32;
 QY 447 LIISFLRCLYEKQKCHIHGEN---KQTSMAAENIGSELPPSATFRFLDM-----L 496
 DB 68 lllvklilwfpqgkcmrhhlqimevmrkqdsrivngngseqlqkeladvimdpmdqpgge 127
 QY 497 KNAKAKSLTES-----LESILSGNKAR-----GLQEHISIVDL-----DSSLSTPLSNT 541
 DB 128 kelvkrslqldgedgplnqlsasltnpvlvlqdkpmsipvkpggdsaspsftpv 187
 QY 542 SKEPSV-----CEKEAL-----PISESSFLL 563
 DB 188 adedsvvfakltylgcasvnaprsevealrmsllrscqqlsdvltlsvpnvsegivrl 247
 QY 564 -----GSSE-DLSSDSESHLPEEPAPL----- 584
 DB 248 dpqntelanypylkllcvrgdhgdpesdcfafteshynaelfrihvfrcelqavsr 307
 QY 585 --SPOQAFRRRANTLSHFPIECQEPPOARGSPGVQRKLMRYHVSSTPTPHERKDFESK 642
 DB 308 lysafafirsa-----kqtplsataaptpdsdlftf-svslel-----kddgk 352
 QY 643 ANHLGSGGTVPKTRHSHR-----QQIFLRVATPQK----- 674
 DB 353 ----gyfsavpkdkdrqcrflrgrldkklviyvqqttnkelaieraicfglllspgkdvrrns 408

QY 675 -----ACDSSRYEDYSELGELPRSP-LEPVCEGDGPGP----- 708
 DB 409 dmhlldlesmgksdgsyvitgswnpkspghfvvneetpkdkvlfmtavdlvitevqe 468
 QY 709 -----PPEEK-----KRTSRE-----LRELWQKA-----ILOQILLRME 738
 DB 469 pvrflletkvrvcspnerlfpwfskrsttenffklkqikqrerkntdtlyevvccese 528
 QY 739 KENQKLO-----ASENDLLNKRKLKLDYEE-----ITPCLKEVTTV 773
 DB 529 sererrkttaspsvrlpgsgsgsvlpspeddeedndepllsgsgdsvskeakilet 588
 QY 774 WEKMLSTPGRSKTKFDMKMHSAVGOGVPRHHRGETWTKFLAEQFHLKHOPPSQPKDVP 833
 DB 589 wgeilis-kwhlnlnvprkqlsslvrgvpealrgewvqllagchndhilev----- 639
 QY 834 YRELLKQLTSQOHAILDGLRFTPTTHPYFSAQLGAGQLSLYNILKAYSLLDQEVGYCQGL 893
 DB 640 yrilickespqdsaitrdinrtfphadyfkdtdggdgslykckaysvyydesigycqgq 699
 QY 894 SFVAGILLHMESEAFKMLKFLMDFMGLRKQYRPMIILQIQMYQLSLRLLHHDYHRDLN 953
 DB 700 sfilaavlllhmpdeeqafsvlvmfdyglrelfkqnfedlhckfyqlerlmqeyipdlyn 759
 QY 954 HLEEHGIGPSLYAAPWELTMEFASORPLGFVARVDFMIFLQGTVEIFKVALSLGSHKPLI 1013
 DB 760 hfidislaehmyasqwlilftakfplynmvfhidlllcegisvifnvalglktakddl 819
 QY 1014 LOHENLETIVDFIKTSLPNLGLVQMEKTTINQVFE-----MDIA-KQLOAYEVEVHYLOEL 1068
 DB 820 ll-tdfegalkkfrvqlpk--tyrseenakkimelacnmkiskqklkkyekyhtnreqg 876
 QY 1069 I-DSSPLSDNQMD-KLEKTNSSLRKQNLLEQLQVANGRTOSLEATIEKLSSSEKUK 1126
 DB 877 aqgedpiferenrrrlqeanrlegenddlahel-----vtskialrkdldnaeakd 930
 QY 1127 QAMLTLELRSALLQVTEELRRRSAPSDREPEC 1160
 DB 931 alnkellmktqkkliaeeekrrleesahlkkmk 964
 RESULT 12
 AAM38650
 ID AAM38650 standard; Protein; 1069 AA.
 XX
 AC AAM38650;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 1795.
 KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;
WPI: 2001-442253/47.
N-PSDB; AAI57806.
Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
Example 3; SEQ ID NO 1795; 10078pp; English.
The invention relates to human nucleic acids (AAI57798-AAI61369) and
the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders.
Note: The sequence data for this patent did not form part of the printed
specification.
Sequence 1069 AA;

Query Match 9.68; Score 580; DB 22; Length 1069;
Best Local Similarity 23.98; Pred. No. 3.7e-37;
Matches 221; Conservative 137; Mismatches 311; Indels 254; Gaps 33;
295 IGQSEVLLISPTDKIALEKFKELSCGQIRHVDHFG-----FICRESSGGGGEHFV 348
191 vsegivrlldqntetanyplkfcvrg-----hdgtpesdcfafteshynaelfr 244
349 CYVQCTNEALVDEIMTLKQAFVAAVQOPAKAPALCEGCPQLSHLKERIEGNNSS 408
245 ihvfceqlqavsrllsfafarsaktpisat-----aptpdsdiftfsvsle----- 297
409 KTKLEQLKHLTLNQEQTIFEEVQKLRPRNEORENELIISFLRC--LYEEKQKEHIHI 466
298 -----ikeddkgvfvav-----pkddrg-----cfklrggldkklviy 332
467 GEMKQTSOMAENIGSELPPSATFRFLDMLKNKAKRSLESILSRGNKARGLQEHSI 526
333 vqqttnkeialerc-----fgl-----llspgkdvrvnsdmhll 365
527 SVDLDSLSLSTLSNTSKEPSYCEKALPISBSFKLLGSSDLSSESHL-----PEE 580
366 --dies-----mgk-----ssdgksyvitgswpks 389
581 PAPLSPOAFRRANTLWSHFPIEQEPPQPPARGSPGVQKLMRYHSVSTTPHERKDFE 640
390 p-----hfqv-----vneetpkdkvlfm 407
641 SKANHLGDSGTPVKTRRHSRQOIFLRVATPQKADSSSRVDEYSELGELPPRSPLEPV 700
408 ttavdl-----vlt-----evqepvrflektvrvcsperl----- 439
701 CEDGPFPPPEKRTSRE-----LRELWQKA-----ILQOILLRMEKENQKLQ----- 745
440 -----fwp-----fskrsttenfklkqkrknntdtlyevvclesesererkrttas 491
746 -----ASENDLLNKLKLDYEE-----ITPCLKEVITVWEKMLSTPGRS 784

Db 492 psvrlpgsgssvippdeedeedndepllsgsgdsvskeackiletwgeilis-kwhl 550
Qy 785 KIKFDMKMHSAVQGVPRHRRGIWFKFLAQFHLKHQFSPKQOPKDVYPYKELLKQLTSQ 844
Db 551 nlrvpqlselsvngypealrgevqlagchndhivke-----yrillitkespq 602
Qy 845 QHAILIDLGRTPPHYFSAQAGQSLYNILKAYSLLDOEVGYCQGLSVFAGILLIHM 904
Db 603 dsaltrdnrtfpahdyfkdgtggdgslykckaysvydeeligcvgqgsflaavllihm 662
Qy 905 SEEEAFKMLFLMDGLMGRKQYRPMIILQIQOMQLSLRLLHDYHRLDYNHLEHEIGPSL 964
Db 663 peeqafsvlvmfdyglrelfkqnfedhckfyqlerlmqeylpdlynhfdlsleahm 722
Qy 965 YAAFWFTMFASQPPFLGFVARVFDIMFLQGTVEVIFKVALSLGSHKPLILQHENLETIVD 1024
Db 723 yasqwlftlftakfplymvfhlidllcegisvfnvalglktkskdlldl-tdfegalk 781
Qy 1025 FIKSTLNLGLVQVEKINQVFE-----MDIA-KOLOAYEVHYHVLQBELI--DSSPLSDNQ 1078
Db 782 ffrvqlpk--ryrseenakkimelacnmkiskgklkkyekyhtmqeqaqgedpferfe 839
Qy 1079 RMD-KLEKNTSSLRKQNLDLLEQLQVANGRIQSLEATTEKLLSESCLKQAMLTLELRS 1137
Db 840 renrlqeammrlqegendldahel-----vtskialrkdldnaeeekadainkelmtkq 893
Qy 1138 ALLQTVELRRRSAPPSDREPEC 1160
Db 894 klidaeeekrrleesaqlkmc 916
RESULT 13
AAB68892
ID AAB68892 standard; Protein: 1069 AA.
XX AC AAB68892;
XX DT 24-APR-2001 (first entry)
XX DE Human RECAP polypeptide, SEQ ID NO: 22.
XX KW Human; RECAP: receptors and associated proteins; cerebrotrophic;
XX KW neurotropic; neuroprotective; anticonvulsant; antiparkinsonian; anti-HIV;
XX KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;
XX KW antithyroid; immunosuppressive; nephrotropic; antitumor; thyromimetic;
XX KW cytostatic; antibacterial; virucide; fungicide; protozoacide;
XX KW antiarteriosclerotic; hepatotropic; gene therapy; infection; cancer.
XX OS Homo sapiens.
XX PN WO200107612-A2.
XX PD 01-FEB-2001.
XX PF 21-JUL-2000; 2000WO-US20035.
XX PR 21-JUL-1999; 99US-0145232.
XX PR 07-OCT-1999; 99US-0158578.
XX PR 12-NOV-1999; 99US-0165192.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Au-Young J, Bandman O, Tang YT, Yue H, Azimzal Y, Burford N;
XX PI Baughn MR, Lu DAM, Hillman JL, Patterson C, Lal P;
XX DR WPI: 2001-168554/17.
XX DR N-PSDB; AAF58616.
XX PT Novel receptors and associated proteins for diagnosis and treatment of
XX PT neurological disorders, immunological disorders including autoimmune/
XX PT inflammatory disorders and cell proliferative disorders such as cancer

Matches 192; Conservative 115; Mismatches 264; Indels 172; Gaps 27;	
Qy	504 LTESLESILSRGNKARGLOEHSISVDLSSLSSTLSNTSKPSCVCEKALPISSESKLL 563
Db	311 lpeavnqvsacfkafqtpypsmcslnsavdmanvts-----dvsgnplntagyeft 364
Qy	564 GS---SEDLSSDESHELPPEAPLSPOQAFRRRANTLSHFPIEQEPP-----OPAR----- 612
Db	365 vsleirervaknsyaavprid-----rgcfkirantdkvecltvktqtpsnvlqlhlerc 418
Qy	613 ----GSPG--VSQR--KLMRYHSVSTETPHERKDPESKANHLGDSGGTPPVKT-----R 657
Db	419 fgvlpagkklvqkdmhldmngyicp-----gggtvatesdsnaq 462
Qy	658 RHWRRQOIFLRVATPQKACDSSSRVEDYSELG-----ELPPRPLEPV----- 700
Db	463 ssswpytiraekaqekafeglnlessktnltvavdlvmrriqepvrfvietptvtigas 522
Qy	701 ---CEDGPGPPPE-----BKRT-----SRELRLWQKAILQOILLRMEK 739
Db	523 enrImdhfmskrpmtlrfylhikrteesnwkvnslpseeite--qpghqgssslkngm 580
Qy	740 EN-----QKIQASENDLLNKLKLDY-----EETPCUKLEVTTVWEKMLS 779
Db	581 nnlsrivrsslasiedcps-----dyssgddepllsrgtgevskdcsgdtldewdpilr 635
Qy	780 TPGRSKIIFDMEK-----MHSVAGQGVPRHHRGEIWKFLAEQFHLKHQPPSQKQKDPYK 835
Db	636 -----ewdsckrpknlapiirvgpealrekiwkla-----nvegrmemnd-kyk 680
Qy	836 ELKOLTQQHALLDLGRTPTTHPYFSAQLGAGOLSLYNILKAYSLLDQEVGYCQGLSF 895
Db	681 lltkctketvlgdrhrtipahckfelgsgsgdalfkskayavndsevgyccglisf 740
Qy	896 VAGILLHMSSEEPKMLKLFMDGLRKQVRPDMIIQIQIQLSRLLDHYHRLDLYNHL 955
Db	741 laasillhmpedafcvlvmaydygldlykagfevlyrlqderlikqqlkhehf 800
Qy	956 EHEHGSLYAAPWFLWFASQPLGFVARVDMIFLQGTVEVFKVALSLGSKHKLILQ 1015
Db	801 tacgiethmasyasqwfiltlytarfpicfvhldvildglpvlfqvavtll-sicesdlr 859
Qy	1016 HENLETIVDFIKSTLPN--LGLVQMEKTIQVFEEMDIAKQAYEVEVHVLEELIDSSP 1073
Db	860 qldfeglikyfrvltlpkcrssqarkvmkqacerki-kkikqyeef----- 906
Qy	1074 LSDNQMDKLEKTNSSLRKQNLDLLEQLQVANGRI-----QSLEATIEKLSSESKLKQAM 1129
Db	907 lllkqhkerlek-----eaqiyenrfgeerrkmqaeidalnqklsakera 952
Qy	1130 LTLELERSALLOTVEELRRSNE 1152
Db	953 vekekktgllqeykqllqrdeq 975
RESULT 15	
AY11605	AA11605 standard; Protein; 94 AA.
XX	AA11605;
AC	
XX	
DT	16-JUN-1999 (first entry)
XX	
DE	Human 5' EST secreted protein SEQ ID NO:257.
XX	
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW	forensic; gene therapy; chromosome mapping; signal peptide;
KW	upstream regulatory sequence; cytokine activity; cell proliferation;
KW	differentiation; haematopoiesis regulation; tissue growth regulation;
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX	thrombolytic; anti-inflammatory; tumour inhibition.
XX	
OS	Homo sapiens.

XX	WO9906439-A2.
PN	
XX	11-FEB-1999.
PD	
XX	
XX	31-JUL-1998; 98WO-IB01233.
PF	
XX	
PR	01-AUG-1997; 97US-0904468.
XX	
PA	(GEST) GENSET.
XX	
XX	Duclert A, Dumas Milne Edwards J, Lacroix B;
DR	WPI; 1999-153700/13.
XX	
DR	N-PSDB; AAX40323.
XX	
PT	New nucleic acids encoding human secreted proteins - obtained from
PT	cDNA libraries derived from liver, lung, large intestine, colon,
PT	thyroid and pancreas tissue
XX	
XX	Claim 27; Page 351-352; 398pp; English.
PS	
XX	AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
CC	human secreted proteins, and encode the proteins given in AAY11533 to
CC	AAY11679, respectively. The proteins given represent the signal peptide
CC	and an N-terminal fragment of a secreted protein. The nucleic acid
CC	sequences can be used for producing secreted human gene products. They
CC	can also be used to develop products for diagnosis and therapy. The
CC	proteins obtained may have cytokine activity, cell
CC	proliferation/differentiation activity, haematopoiesis regulating
CC	activity, tissue growth regulating activity, reproductive hormone
CC	regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC	thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC	activity, tumour inhibition activity or other activities. The products
CC	can be used in forensic, gene therapy and chromosome mapping procedures.
CC	The sequences can also be used for obtaining corresponding promoter
CC	sequences. The nucleic acids encoding the signal peptide can be used for
CC	directing extracellular secretion of a polypeptide or the insertion of a
CC	polypeptide into a membrane, or importing a polypeptide into a cell.
XX	
SQ	Sequence 94 AA;

Query Match 7.9%; Score 481; DB 20; Length 94;
Best Local Similarity 97.9%; Pred. No. 7.3e-31;
Matches 92; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1 MEPIFTARKHLLPNEVSVDFGLQLVGLSLPYHSLTTPMLPWVVAEVRRLRSQSTRKEPV 60
Db	1 mepitfarkhllpnevsvdfglqlvgsalpvhsllttmpmpwvvaevrrlrsqstrkepv 60
Qy	61 TKQVRLCVSPSGLRCPEPGRSQWDPLIYSSIF 94
Db	61 txqrlcvspsglrcpepgrsqwdpliyasif 94

Search completed: August 28, 2002, 16:23:49
Job time: 82 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 28, 2002, 16:22:27 ; Search time 16.95 Seconds
(without alignments)
1683.132 Million cell updates/sec

Title: US-09-762-311-5
Perfect score: 6055
Sequence: 1 MEPTTFARKHLLPNEVSVD.....RSAPSDREPECTQPEPTGD 1168

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents:AA:*
1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2.6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2.6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4976	82.2	1141	1	US-08-363-300-2
2	232.5	3.8	3248	1	Sequence 2, Appli
3	232.5	3.8	3248	5	Sequence 1, Appli
4	211	3.5	2482	1	Sequence 6, Appli
5	197	3.3	376	1	Sequence 33, Appli
6	196.5	3.2	1093	5	Sequence 1, Appli
7	190.5	3.1	885	2	Sequence 4, Appli
8	190.5	3.1	885	2	Sequence 4, Appli
9	180	3.0	1886	4	Sequence 3, Appli
10	175	2.9	1939	4	Sequence 1, Appli
11	169	2.8	816	2	Sequence 6, Appli
12	169	2.8	816	2	Sequence 6, Appli
13	160.5	2.7	1312	2	Sequence 51, Appli
14	159.5	2.6	1312	2	Sequence 148, App
15	156.5	2.6	1001	4	Sequence 2, Appli
16	153	2.5	2101	1	Sequence 4, Appli
17	153	2.5	2101	1	Sequence 4, Appli
18	153	2.5	2101	1	Sequence 4, Appli
19	153	2.5	2101	1	Sequence 4, Appli
20	153	2.5	2101	2	Sequence 4, Appli
21	153	2.5	2101	4	Sequence 1, Appli
22	153	2.5	2101	5	Sequence 4, Appli
23	151.5	2.5	835	2	Sequence 4, Appli
24	151.5	2.5	1388	4	Sequence 2, Appli
25	150	2.5	1164	4	Sequence 2, Appli
26	148.5	2.5	1354	3	Sequence 2, Appli
27	142.5	2.4	1104	4	Sequence 4, Appli

28 141.5 2.3 2548 4 US-09-172-422-1 Sequence 1, Appli
29 138 2.3 988 2 US-08-286-819A-19 Sequence 19, Appli
30 138 2.3 988 3 US-08-980-357-19 Sequence 19, Appli
31 136.5 2.3 1122 2 US-08-619-198-3 Sequence 3, Appli
32 136.5 2.3 1389 2 US-08-619-198-5 Sequence 5, Appli
33 135 2.2 557 4 US-08-979-608A-5 Sequence 5, Appli
34 134 2.2 1618 1 US-07-853-913-4 Sequence 2, Appli
35 132.5 2.2 710 4 US-09-079-812E-2 Sequence 4, Appli
36 132 2.2 2485 4 US-09-290-640-46 Sequence 46, Appli
37 131 2.2 976 4 US-09-722-139-2 Sequence 2, Appli
38 130.5 2.2 1375 4 US-09-722-139-2 Sequence 2, Appli
39 130.5 2.2 1863 2 US-08-603-753D-2 Sequence 2, Appli
40 130.5 2.2 1863 4 US-09-099-753-2 Sequence 2, Appli
41 130.5 2.2 1863 4 US-09-986-106-2 Sequence 2, Appli
42 130.5 2.2 1863 4 US-09-007-678B-49 Sequence 49, Appli
43 130.5 2.2 1898 1 US-08-056-200-94 Sequence 94, Appli
44 130.5 2.2 1898 2 US-08-800-644-94 Sequence 9, Appli
45 130.5 2.2 1976 3 US-09-024-020B-9 Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-363-300-2
; Sequence 2, Application US/08363300
; Patent No. 5700927
; GENERAL INFORMATION:
; APPLICANT: Zon, Leonard and Richardson, Paul
; TITLE OF INVENTION: Tbc1 Gene and Uses Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,300
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04590/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1141 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-363-300-2

Query Match 82.2%; Score 4976; DB 1; Length 1141;
Best Local Similarity 87.9%; Pred. No. 0;
Matches 970; Conservative 47; Mismatches 74; Indels 12; Gaps 5;

QY 37 MPMLPWWAEVRLSRQSTRKEPTVKQVRLCVSPSGRLCEPGRSQOWDPLIYSIFEC 96
Db 1 MPMLPWWAEVRLSRQSTRKEPTVKQVRLCVSPSGRLCEPGRSQOWDPLIYSIFEC 60
QY 97 KPORVHKLHNSHDPGVFACLIKEDAVHROSICVYFKADQTKVPELISSIRQAGKIRQ 156

```

Db 61 KPOVHKLHNSHDPVSFACLIKEDAAHQSLCYVFKADDQTKVPEIISIRQAGKIARQ 120
QY 157 BELHCPSEDDTSKFEVLFQGRVVAHKKAPPALIDECIEKFNHVSGRGSESRPNP 216
Db 121 EELRCPSEDDTFAKFEVLFQGRVVAHKKAPPALIDECIEKFNHVSGRGRTD----- 174
QY 217 PHAAPTGS-QEPPRRPKRSFSPQGLRSALAFRAKELDGGRLSSGFTSSPEESDIENHLIS 275
Db 175 -WEAPTQPSAPGPRPKRSFSPQGLRSALAFRAKELDGGRLSS-TFSSP-DNDIENHLIG 231
QY 276 GHNVPTDIEENRTMLFTTIGSEVYLIISPDTKKIALEKNFKIEISCSGIRHVDHFGFI 335
Db 232 GHNVPTDMEENRTMLFTTIGSEVYLIISPDTKKIALEKNFKIEISCSGIRHVDHFGFI 291
QY 336 CRESSGG--GGFHVCYVFOCTNEALVDEIMTLTKOAFVAAVQOAKAPALCEGCPLO 393
Db 292 CRESSGGGGGFFHVCYVFOCTNEALVDEIMTLTKOAFVAAVQOAKAPALCEGCPLO 351
QY 394 SLHKLCEIRIEGMSSSKTKLEQLKHLTLTNOEQATIFEVQKLRPRNEORENELIISFLR 453
Db 352 GLHKLCEIRIEGMSSSKTKLEQLKHLTLTNOEQATIFEVQKLRPRNEORENELIISFLR 411
QY 454 CLYEKQKEHIIHGMKQTSQMAENIGSELPPSARFRFLDMLKNKAKRSLTESLESILS 513
Db 412 CLYEKQKEHSHTGAPKQTLQVAAENIGSDLPSPASRFRDLNKNRAKRSLSLESILS 471
QY 514 RGNKARGLOEHSISVDLSSLSSTLNTSKEPSVCEKEALPISESSFKLLGSSDLSDDS 573
Db 472 RGNKARGLOHSASVDLSSSTLNTSKELSMGKEAPVSETFKLLGSSDLSDDS 531
QY 574 ESHLPPEPAPLSPQAFRRRANTLSHFFIECQPPPPARGSPGVSKLMRYHSVSTETP 633
Db 532 EGHIAESALLSPQAFRRRANTLSHFFIECQPPPPARGSPGVSKLMRYHSVSTETP 591
QY 634 HERKDFESKANHLGDSGGTPVKTRRHSWROQIFLRVATPQKACDSSSRVEDYSELGELPP 693
Db 592 HERKDFESKANHLGDSGGTPVKTRRHSWROQIFLRVATPQKACDSSSRVEDYSELGELPP 651
QY 694 RSLPEVCEGDPGPPPEEKKRTSRELRLWOKAILQOILLLRWKENOKLOASENDLLN 753
Db 652 RSLPEVCEGDPGPPGQYKRRHSHAFSECGKRPSCSCLVRWKENOKLOASENDLLN 711
QY 754 KRLKLDYEITPCLKEVTTWEEKMLSTPGRSKIKFDMKMSHSAVGQVPRHRHGEIWKFL 813
Db 712 KRLKLDYEITPCLKEVTTWEEKMLSTPGRSKIKFDMKMSHSAVGQVPRHRHGEIWKFL 771
QY 814 AEOFHLKHQPPSKQPKDVPYKELLKQLTSQOHAILLDGRTPPTTHPYFSAQLGAGQLSL 873
Db 772 AEOFHLKHPPSPKQPKDVPYKELLKQLTSQOHAILLDGRTPPTTHPYFSAQLGAGQLSL 831
QY 874 YNLIKAYSLLDQEVGYCOGLSFVAGILLHMSBEEAFKMLKFLMDGLRKQYRPMIIL 933
Db 832 YNLIKAYSLLDQEVGYCOGLSFVAGILLHMSBEEAFKMLKFLMDGLRKQYRPMIIL 891
QY 934 QIQMYQLSRLLDHYHRLDYNHLEBEHETGSLYAAPWFLTMFASQFPLGFRVAFVDFMIFLQ 993
Db 892 QIQMYQLSRLLDHYHRLDYNHLEBEHETGTPPTYAAPWFLTMFASQFPLGFRVAFVDFMIFLQ 951
QY 994 GTEVIFKVALSLGSHKPLILOHENLETIYDFIKSTLPNLGLVQMEKTIINQFEMDIQAK 1053
Db 952 GSEVIFKVALSLGSHKPLILOHENLETIYDFIKSTLPNLGLVQMEKTIISQVFEEMDIQAK 1011
QY 1054 LQAVEYHVLQELIDSSPLSDNORMDKLEKTNSSLRKONLDLLELOQVANGRIOSLEA 1113
Db 1012 LQAVEYHVVQELIESPLSDNORMDKLEKTNSSLRKONLDLLELOQVANGRIOSLEA 1071
QY 1114 TIEKLSSSESKLQAMLTLELER 1136
Db 1072 TVEKLLTSESKLQALTLLEVER 1094

```

RESULT 2
US-08-353-700-1

```

; Sequence 1, Application US/08353700
; Patent No. 559919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TRANSLANTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 3.8%; Score 232.5; DB 1; Length 3248;
Best Local Similarity 19.3%; Pred. No. 2e-09;
Matches 240; Conservative 191; Mismatches 440; Indels 373; Gaps 55;

QY 74 RCPEPGRSQWDPDIYSSIFECK-PORVHKLHNSHDP-S-YFACLIKEDAVHRSICVY 131
Db 120 RCKSELSQQAASADVSLNCPNTQ---KFTTPTLTPSYSGSKYEDLKEK-----YN 172
QY 132 FKADDTKVPETIISIRQAGKIARQBELHCPSEDDTFSKFEVLFQGRVTVAAHKAPPA 191
Db 173 KEVEERKRLEAEYKALQ-----AKK-----ASQTLQA 200
QY 192 LIDECIEKFNHVSGRGSESPRNPPHAAPTGQEPVPRPMKRSFQGLRSALAFRAKEL 251
Db 201 TMHRIARHQASSVFSWQEQETPSHLSSNSQRTPIRRDFSASYFSGELEVPSRSTLQ 260
QY 252 DGGI-RSSGFFS-----FEESDIENHL-----ISGHNVPTDIEENRTMLF 293
Db 261 IGRDANSFFGNSSPHLLDQLKAQNOELRNKINELELRLOGHEKMKQVANKFBLQL 320
QY 294 TICQSEVYLISPD-----TTKIALEKNFKIEISCSGIRHVDHF 332
Db 321 QLEKAKVELIEKEKVLNKRDELVRTTTAQQDOASTKTYALEQKLKLT-----E 369
QY 333 GFICRESSGGGFFHVCYVFOCTNEALVDE-----IMTLTKOAFVAAVQ 378
Db 370 DLSQCRQNAESA-----RCSLEQKIKEKEFEQELSQRQSFQTLQO-----ECIQM 417

```


Qy	579	E-----	EPAPISPOQAQRRRRANTLUSHPPICEQPPQPARGSPG	616	
Db	1183	EVOLMTKIEACIBLEKIVGELKENSJLSEK-----	LEYTSCDHQELLQORVETSEG	1233	
Qy	617	VSQKLMRYHSVSTETPHERKDPFESKANHLGDSGGTVPKTRRHRSWROIF-----		666	
Db	1234	LNSDLEM-----HADK-----	VAKYNDSWKERFLDVENELSRIR	1276	
Qy	667	-----LRVATPQKAC-----	DSSSRYEDYSELGELPPRSPLEVPCEGPPG	707	
Db	1277	SEKASIEHEALYLEADLEVYVQTEKLCLEKDNEKQVI-----	VCLBEELS	1322	
Qy	708	PPPEEKRTSRBELRWQRAILOQILLRMKXENOKLOASENDLLN-----	K	754	
Db	1323	VYTSERNQLRGELDTMSKKTATDOLSEKMKKEKTEHOSSECLHICIQVAEAEVKEKTE	1382		
Qy	755	RLKLDYEETPCLEKVEYTWKEMWLSPTGSRSKIKPFOMEKMHSAVGQVPRHRHGEIWKFLA	814		
Db	1383	LLQTLSDVSVELLKDKTHQELQKS-----	LEKDSQALS-----	L 1417	
Qy	815	EQFHLKHOPFSKOQPDVPYKE---LLKOLTQQOH-----	AILEDLGRTPPTHY	861	
Db	1418	TKCELENOIAQLNKEKELLVKESESLOARLSSDDEYKLVNKSALFAALVEKGE-----	1470		
Qy	862	FSQAQAGAGQSLYNIUKAYSLL-----		883	
Db	1471	FALRLSTQBEVHQLRGARGIEKLRVIRTEADKKQLHIAEKLRERENDSLDKVENLERE	1530		
Qy	884	-----DOEYGVCGGLSFVAGILLLHMSBEEAFKMLFLMFDM-----	GLRQYRPD	929	
Db	1531	LQWSEENQELVILDAGNSKAQVETLKTQTEEMARSLKVFELDLYLRSEKENLTKQLOEK	1590		
Qy	930	MIILQIOMYQLSRLHDYRDLYNHLHEHE---IGPSLYAAPWFLTMFASQPPLGFVARVF	987		
Db	1591	-----QGGLSELDKLLSSF-KSLEEKEQAEIQIKEESKTAVEMLQNQLKELNEAVALCG	1645		
Qy	988	DMTFLOQTE-----	VIFKVALSILGSHRP---	LILQHENLFTIVDFTKST	1029
Db	1646	DOETMKATEOSLDPPIDEEHQLRNLNSTEKLRLARLEADEKKQCLVCVLOQLKESESHADLLKGR	1705		
Qy	1030	LPNL-----GLVQMEKTIQVFM-----	DIAKOLQAYEVEYHVL---ORELI	1069	
Db	1706	VENLERELETARTNQBEHAALENBSNGEVETLKAKIEGTOSLRGLEQUDVVYTIKSEKNL	1765		
Qy	1070	DSSPLSDNQMDKLEKTNSS-----	LRKONLDLLEQLQV----	ANGRIQSL	1111
Db	1766	TNELQKEQBRISELEINTINSFENILQKEQEKVQMKESKSTAMEMLOQQLAKELNERVAAL	1825		
Qy	1112	EATIEKLLSSSEKQKAMITLELERSALLQTVVEULR	1147		
Db	1826	HNDQEAQKAEQNLSQVCLELEKRAQLLQGLDEAK	1861		

```

RESULT      5
US-08-253-155A-33
: Sequence 33, Application US/08253155A
: Patent No. 5691147
:
: GENERAL INFORMATION:
:   APPLICANT: Gyuris, Jeno
:   APPLICANT: Drietta, Giulio
:   TITLE OF INVENTION: CDK4 Binding
:   NUMBER OF SEQUENCES: 95
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: LAHIVE & COCKFIELD
:   STREET: 60 State Street
:   CITY: Boston
:   STATE: MA
:   COUNTRY: USA
:   ZIP: 02109
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: ASCII(text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/253,155A
: FILING DATE: 02-JUN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36,709
: REFERENCE/DOCKET NUMBER: MII-028
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 376 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FRAGMENT TYPE: N-terminal
: US-08-253-155A-33

Query Match 3.3%; Score 197; DB 1; Length 376;
Best Local Similarity 23.9%; Pred. No. 5.5e-08;
Matches 66; Conservative 41; Mismatches 99; Indels 70; Gaps

Qy 771 TTWVEKMLSPGRSKIFEDMEKHSAVGQGVPRHHRGEIWKFLAEQFHLKHQFPSPKQPK 830
Db 774 TSKYMEML---GEWETKYSKSLDRIVYKIGIPNIRGVPWVSLNTQEIKNPNRIQ-- 128
Qy 831 DVPYKELLKOLTSSQAHAILDLGRTFTPTHPYFSAQAGAGQSLYNILTKAYSLDDQEVGYC 890
Db 129 --IMKERGRSSSEHIIHIDLVRTTLRNHVFFRDYRGAQRELFYILLAYSEYNPEVGYC 186
Qy 891 QGLSFVAGILLHMSSEEAQKMLKFLMFD-----MGLRKQYRPDMIIQ 934
Db 187 RDLSHITALFLLYPEEDAFWALVQLLASERHSLPGFHPNGTVOGLDQ----- 237
Qy 935 IQWYQLSRLLHDYHRDLYNHLGEIGSPSLAAPW-----FLTMFASQF 978
Db 238 -----QEHVPKSPQKTNWQHDKEGLCGQCASLGCLLRLNLDIG 276
Qy 979 PLGEVARVDFMIFLQGTGVIFKV-ALSLGLSHKPLI 1013
Db 277 SLGTLRLDDVLYEGEOVLMPITTSIALKYOORLKM 312

```

```

6
RESULT
PCT-US93-03077-1
; Sequence 1, Application PC/TUS9303077
; GENERAL INFORMATION:
; APPLICANT: Board of Regents, The University of Texas System
; APPLICANT: Gaynor, Richard B.
; APPLICANT: Wu, Foon Kin
; TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
; TITLE OF INVENTION: REGULATING GENE EXPRESSION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03077
; FILING DATE: 19930331
; CLASSIFICATION:
;

```


Db 507 DAKLRLEV-----NMQALGQFERDQARDEQNEEKRRQQLQQLHETEL-----EDERN 557
Qy 862 FSAQLGAGOLSYNLIKAYSLDDQEVGYCGLSFVAGILLHMSBEEAFKMLKFLMFMG 921
Db 558 ERALAAAARKKLEGLDKLEL-----QADSAIKG-----REEAIKQ-----593
Qy 922 LRKQYRPMIILQIQMYQLSRLLDHYR-----DLYNHLEBE-IGPSLYAAPWFLTMFASQ 977
Db 594 LRK-----LOAQMKDFORELEBARASDEIFATAKENKAKSLEA-----DIMQLQ 640
Qy 978 FPLGFVARYDFMIFLOGTEVIFKVALSLGSHKPLILOHE--NLETIVDFIKSTLPNLGL 1035
Db 641 EDLAAARARKQADLEKELELASSLSGRN---ALQDEKRRLEA-----RI 685
Qy 1036 VQMEKTIQVFMNDIAKQLQAYEYHVHLELIDSSPLSDNQRMDKLEKTNSSLRKQML 1095
Db 686 AQLEEELE---BEQGMNEMASDRVRKATQAOQLSNELATERSTAQKNESARQQLERQNK 742
Qy 1096 DLLEQLQVANGRIQS-----LEATI-----EKLSSSESKLQOAM 1129
Db 743 ELRSKLHEGAVKSKFTIAALEAKIAQLEEQVEQAREKQATKSLKQDKKLKLEIL 802
Qy 1130 LTLLELRSALLQVEELRRRSAPSDREPECTQPE 1164
Db 803 LOVEDERKMAEQYKEQAEKGNARVKQKQLEEA 837

RESULT 9

US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Leinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,105
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crook, Wanneil M.
; REGISTRATION NUMBER: 31,071
; REFERENCE/DOCKET NUMBER: 3595-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1886 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-938-105-3

Query Match 3-0%; Score 180; DB 4; Length 1886;
Best Local Similarity 18.7%; Pred. No. 1.7e-05;
Matches 161; Conservative 135; Mismatches 320; Indels 246; Gaps 33;

Qy 422 TNOEQATIFEVQKLR---PRNEORENEL---IISFLCLYBEKKEKHIHIGEMKOTSOM 475
Db 794 TEKEMANKKEEGRVKDALESEARKKELEKMSVLL-----QEKNDLQLOV-QAQODNLA 848
Qy 476 AENIGSELPPSATRFLD-MLKNKAK-----RSUTESLESILSRG-----NKARGLOEH- 524
Db 849 DAE-----RCQLIKKIQLEAKVKEMTERLEDEEMNAELTAKRKLDEBC 896
Qy 525 -SISVDLSSLSSTLSNTSKPSVCEKALPISESFFKLLGSSEDLSSPSESHLPEPAP 583
Db 897 SELKKDID-DLELTAKVEKEKHATENKVNITEEMAGLEDEIIAKLTREKKAQLEAHQA 955
Qy 584 LSPQAFRRRANTLSHFPIECQPPQARGSPGVQSKMLRY-----625
Db 956 LDDLAQAEEDKVTNLKSKVKLEQQVDDLEGS--LEQEKVRMDLERAKRKLKGLKLTQE 1013
Qy 626 -----HSVTETTERKDPESKANHLGDSGTPVKTRRHSWRQOIFLRVATPPQACDS 678
Db 1014 STMDLENDKLOLEEKKKKEFD-----ISOQNSKIEDEQALQLOKKLKEN 1060
Qy 679 SSRYEDYSSELGELPPRSPLEPCVEDGPGPPPEKKRTSRELRLWQKAILQOILLRME 738
Db 1061 QARIELEE--EL-----EARTARAKVEK-----LRSDLTRELE 1093
Qy 739 KENQKIQ-----ASENDLNLKRLKLDYEBITPCLEKVEVTVWEKMLSTPGRSKIKFDMKM 793
Db 1094 EISERLEEAGGATSVQIENMKKREAEFQKMRDRLEATL-----Q 1133
Qy 794 HSAVGQVPRHRGELWFLAEQFHLKHQFPKQKQDPVPYKELLKQLTSSQOHA- 848
Db 1134 HEATAAALRKKHADSVAE--LGEQIDNLQVRVKQKLEKSEFKLEDDVTSHEQIIKAKA 1192
Qy 849 -LIDLGRTPFTHP-YFSAQLGAGOLSYNLIKAYSLDDQEVGYCGLSFVAGILLHMS 906
Db 1193 NLEKVSRTLEDOANEYRVKLEEAQRLNDFTTORAKLTE-----NGELARQLEE 1242
Qy 907 EEA-----FKMLKFLMDMG-----LRKOYRPM- 930
Db 1243 KEALIWOLTRGKLSVTQOMEDLKROLEEBGKAKNALAHALQSAHRHDCDLLREQYEBEEMA 1302
Qy 931 -----IILQIQMYQLSRLLDHYRDLNLYNLEHEIGPSLYAAPWFLTMFASQPLGFVAR 985
Db 1303 KAEQRLVLSKANSEVAQWRTKYETDAIQRTTELEBAKKLAQR-----1345
Qy 986 VEDMIFLOGTEVIFKVALSLGS-----HKPLILHENLETIVDFIKSTLPNLGLVQMEK 1040
Db 1346 -----LQDAEAEVAVNAKSSLEKTKR---LQNEIEDLVQDVERSNAAAAALDKKOR 1396
Qy 1041 TINQVFMEDIAKOLQAYEYHVHLELIDSSPLSDN-----ORMDKLEKTN 1088
Db 1397 NFDKI---LAEWKQKYEESQSELESSQKARSLSLTELFLKNAYEESLEHLETFKRENK 1452
Qy 1089 SLRKONLDLLEQLQVANGRIQSLIETIEKL-----LSSESKLQOAMLTLEERSALL- 1140
Db 1453 NLQEEISDLTEQLEGEGKNVHELEKIRKOLEVEKLEQSALEEAASLEHEGKILRAOL 1512
Qy 1141 ---QVVEELRRRSARPSDREPE 1159
Db 1513 EFNQIKAEIERKLAE-KDEEME 1533

RESULT 10

US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12

NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match 2.9% Score 175; DB 4; Length 1939;
Best Local Similarity 18.1% Pred. No. 4.4e-05;
Matches 150; Conservative 139; Mismatches 358; Indels 184; Gaps 27;
Qy 422 TNOQATIFEVQKLR---PNEORENELIISFLRCLYEK-----QKEHIHGENKQT 472
Db 847 TEKEMATKKEFGRIKTEKSEARRKELEKMYSLQEKNDLQVQAEQDNDAER 906
Qy 473 SOMAAENIGSELPPSATFRDMLNKAARSLTESILSRGNKARGLOEH--SISVDL 530
Db 907 CDQLIKN-----KIQLEAKVKEMNERLEDEENNAELTAKKRKLEDECSLKKDI 956
Qy 531 DSSLSSTLSNTSKPSVCEKALPISSESKLLGSSDLSDSSESHLPEEPAPISPOQAF 590
Db 957 D-DLEATLAKVEKKEKATENKVNKLTEEMAGLDELIATKLTREKALQEAHQALDDQLVE 1015
Qy 591 RRRANTLSHFPIEQEPPOPARSGPGVYSORKLMRY----- 625
Db 1016 EDKVNLSKSKVKLEQQVDDLEGS--LEQEKVVRMDLERAKRKLEGDLKLTQESIMDLEN 1073
Qy 626 HSVTETPHERKDFESKANHLGSGGTPVKTRRHSWROQIFLRVATPKACDSSSRVEDY 685
Db 1074 DKQLLEKLLKKEFD-----INQNSKIEQALALOKKLEKNAARTEEL 1120
Qy 686 SE--LGELPPRSPLEPCVEDGPPPPPEKKRTSRRELWOKAILQOILLRMEKENQK 743
Db 1121 EEELEAERTARAKVEKLRS-----LSRELEEISE-----RLEEAGG- 1157
Qy 744 LQASENDLLNRLKLDVEETPCLEKVTWWEKMLSTPGRSKIFDKMEKHSVAGQGVPR 803
Db 1158 -ATSVQIEMNKKRAEFQKMRDLDEATL-----QHEATAAALRK 1196
Qy 804 HHRGEIWKFLAEQFHLKHQPPSKQOPKVPYKELLKOLTSOQHA1-----LIDLGRTP 857
Db 1197 KHAQSVAE-LGEQIDNLRQVKQKLEKSEPKLEDDVTSNWEQIRAKANLEKVSRTLE 1255
Qy 858 THP-YFSAQLGAGQSLNLIKAYSLLDQEVGYCOGLSFVAGIILLHMSSEAF----- 910
Db 1256 DQANEYRVKLEEAQRSLNDFTTORAKLQTE-----NGELARQLEEKALISQLTR 1305
Qy 911 KWLAFMFDMLGRKQYRPFDMILIQIOMYOLSRLLHDYH--RDLYNHLFE--HEIGPSIYA 966
Db 1306 GKLSYTOQMEDLKRQLEEEGAKALAHALQASARHDCOLLREQYEETEAKAELQVLSK 1365
Qy 967 APFWLTFASQFPLGFVARVDFMIFLQTEVIFKVALSLGSHKPL----- 1012
Db 1366 ANSEVAQRWRTYETDAIQRTTEL-----EAKKKLAQRLQDAEEAVEAVNAKCSLEKTK 1420
Qy 1013 -ILQHENLETVDFTKSLPNIQVMEKTTINQVEMDIQALQAYEVEYHVLOEELIDS 1071
Db 1421 HRLQNEIEDLVMDVRSNAAAALDKQKQNFDK1----LAEWKQKYEESQSELESSKEA 1476
Qy 1072 SPLSDN-----QRMDKLEKTNSSIRKQNLIDLEOLQVANGRIQSLIEATKEL- 1118
Db 1477 RSLSTELPKLNAYEESLEHLETFKRENKNLQEEISLDTEQLGEGGKNVHELEKVRKOLE 1536
Qy 1119 ---LSSEKLLKQAMLTLEERSALL-----QTVEEELRRRSAPPSDREPE 1159
Db 1537 VEKLEQLQSALEAEASLEHESGKILRAQLEFNQIKAEIERKLAE-KDEEME 1586

RESULT 11
US-08-533-306A-6
; Sequence 6, Application US/08533306A

Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 2.8% Score 169; DB 2; Length 816;
Best Local Similarity 19.9% Pred. No. 3.6e-05;
Matches 170; Conservative 112; Mismatches 321; Indels 250; Gaps 34;

Qy 381 KAPQALCEGCPQLSHLKCERIEGMSKTKLEKQKHLTTLTNEQATIFPEVQKLRPN 440
Db 97 KAPMILNGVCYVIWKGWIDLQRLDGMGCLFEDEERAQOEDALAQQ-----AFEEA 147
Qy 441 EORENELIISFLRCLYEKQKEHIHGENKQTSQMAAENIGSELPPSATFRDMLKNKA 500
Db 148 RTRE-----FEDRDRSH----- 159
Qy 501 KRSITESLESILSRGNKARGLOEHSISVDL--SSLSTLSNTSKPSVCEKALPISSESS 559
Db 160 REEMENEVESVTGMLNEAEG---KAIKLAKDVASLSLOLQDTQE---LLOEETROKLNVS 213
Qy 560 FKLLGSSDLSDSSESHLPEEPAPLSVQOAFRRRANTLSHFPIECQEPPOPARSGVSYQ 619
Db 214 TKLQLEEEERNS-LQDQLDEE---WEAKQNLEHISTLN---IQLSD-----SK 255
Qy 620 RKLARYHSVSTETPHERKDFESKANHLGDSGGTPV-----KTRRHSWRO----- 663
Db 256 KKLQDFASTVLEALBEGKKRFQKEIENLTQOYEEKAAAYDKLEKTKNRLQOELDLVDLD 315
Qy 664 ---QIFLRVATPKACDS-----SSRYEDYSELGELPPRSPLEPCVEDGPGFPPPEE 712
Db 316 NORQLVSNLEKKQKQKFDQLLAEKKNISSKYAD----- 348
Qy 713 KKRTSRELRLWQKAI-----LQOILLRL--WEKENQKLOASENDLLNKR--LKLDYEEI 763
Db 349 RDRAEAAREKETKALSARALEALEAKEELERTNKMKAEMEDLVSSKDDVQKGNHEL 408

QY 764 TPCLKEVTTWVKMLSTPGRSKIKFDMKMHSAVGQGVPRHHRGEIWKFLAEQFHLKHQF 823
 Db 409 EKSKRALETQMEEM-----KTQLEEELEDELQASEDAKLREV-----NMQALKGQF 454
 QY 824 PSKQPKDVPYKELLKQTSQQHAILIDLGRTPTTPHYFSAQAGQSLSYNLKAYSLL 883
 Db 455 ERDLQARDEQNEEKRRQLQRLQHEYTEL-----EDERNERALAAAKKLEGDLKDEL- 509
 QY 884 DQEVGYCOGLSFVAGILLHMSSEEAFAKMLKFLMDFMGLRKQYRPMIILQIQMYQLSRL 943
 Db 510 -----QADSAIKG-----REEAIKQ-----LRK-----LQAMKDFQRE 538
 QY 944 LHDYHR---DLYNHLBEEH-IGPSLYAAPWFLTMFASQFPLGFVARVDFMIFLQGTGVIF 999
 Db 539 LEDARASRDEIFATAKENEKAKSLEA-----DLMLQLEDLAAAEARKQADLEKEELAE 593
 QY 1000 KVALSLGSHKPLILQHE---NLETIVDFIKSTLPNLGLVQMEKTIQVFMEDIAKQLQAY 1057
 Db 594 ELASSLSGRN---ALQDEKRRLEA-----RIALQEELE---EEQGNMEAMSD 635
 QY 1058 EYEVHYVLOEELIDSSPLSDNQRMDKLEKTNSSLRKQNLDLLEQLVANGRIQS----- 1110
 Db 636 RVKATQQAQAEQLSNELATERSTAKNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIA 695
 QY 1111 -LEATI-----EKLSSSEKSKQAMLTLELERSALLQTVVEELRRSA 1151
 Db 696 ALEAKTAQLEEQVEQAREKQAATKSLKQDKKLLKEILLQVEDERKMAEQYKQAEKGNA 755
 QY 1152 EPSDRPECTOPE 1164
 Db 756 RVKQLKRLQEEAE 768

RESULT 12

US-08-742-923A-6
 ; Sequence 6, Application US/08742923A
 ; Patent No. 5869611
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Pu
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Siciliano, Michael J.
 ; APPLICANT: Claxton, David
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16
 ; TITLE OF INVENTION: Rearrangements
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/742,923A
 ; FILING DATE: NO. 5869611 member 1, 1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Deann F.
 ; REGISTRATION NUMBER: 36683
 ; REFERENCE/DOCKET NUMBER: 12115-00869DVC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 641-1600
 ; TELEFAX: (810) 641-0270
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 816 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear

MOLECULE TYPE: protein
 US-08-742-923A-6

Query Match 2.8%; Score 169; DB 2; Length 816;
 Best Local Similarity 19.9%; Pred. No. 3.6e-05;
 Matches 170; Conservative 112; Mismatches 321; Indels 250; Gaps 34;

QY 381 KAPAQICEGPCLOSLLKLCERIEGMNSSKTKLEQKHLTLTNOEQATIFEYQKLRPN 440
 Db 97 KAPMIUNGVCVWIKWIDQLQDGMGCLFEDEERAQOEDALAQQ-----AFEAA-----RR 147
 QY 441 EORENELIISFLCLCYEEKQKEHIHIGEMKQTSQMAAENIGSELPPSATFRFLMLNKA 500
 Db 148 RYRE-----FEDRRSH----- 159
 QY 501 KRSLTSLSEIISGRNKAQGLQHSISVDLD--SSLSTSLNTSKPSVCKEKALPISSE 559
 Db 160 REEMENEVESVTGMLNEAEG---KAIKAKDVASLSQLODQOE---LLOEETRQKLNVS 213
 QY 560 FKLLGSSSEDLSSDSHLPPEAPLSPQQAFFRRANTLSHFPIECOEPPQARGSPGVQ 619
 Db 214 TKLQRLQEEERNS-LQDQDEE---MEAKONLERHISTLN---IQLSD-----SK 255
 QY 620 RKLRYHVSYSTETPHERKDFESKANHLGDSGGTPV-----KTRRHSWRQ----- 663
 Db 256 KKLQDFASTVEALEEGKRFQKEIENLTQYEEKAAAYDKLEKTKNRLQOQLDVLVDLD 315
 QY 664 ---QIFLRVATPOKACDS-----SSRYEDYSELGELPPRSPLEPVCEDGGFGPPPEE 712
 Db 316 NQRLVSNLEKKQKQKFDQLLAEEKNISKYAD-----E 348
 QY 713 KRTSRELRLWQKAI-----LQOILLR--MEKENQKLOASENDLLNKR--LKL DYEI 763
 Db 349 RDRAEAREKETKALSARALEEAELEKELEKTKNMLKAEEMEDLVSSKDDYGVKNVHEL 408
 QY 764 TPCLKEVTTWVKMLSTPGRSKIKFDMKMHSAVGQGVPRHHRGEIWKFLAEQFHLKHQF 823
 Db 409 EKSKRALETQMEEM-----KTQLEEELEDELQASEDAKLREV-----NMQALKGQF 454
 QY 824 PSKQPKDVPYKELLKQTSQQHAILIDLGRTPTTPHYFSAQAGQSLSYNLKAYSLL 883
 Db 455 ERDLQARDEQNEEKRRQLQRLQHEYTEL-----EDERNERALAAAKKLEGDLKDEL- 509
 QY 884 DQEVGYCOGLSFVAGILLHMSSEEAFAKMLKFLMDFMGLRKQYRPMIILQIQMYQLSRL 943
 Db 510 -----QADSAIKG-----REEAIKQ-----LRK-----LQAMKDFQRE 538
 QY 944 LHDYHR---DLYNHLBEEH-IGPSLYAAPWFLTMFASQFPLGFVARVDFMIFLQGTGVIF 999
 Db 539 LEDARASRDEIFATAKENEKAKSLEA-----DLMLQLEDLAAAEARKQADLEKEELAE 593
 QY 1000 KVALSLGSHKPLILQHE---NLETIVDFIKSTLPNLGLVQMEKTIQVFMEDIAKQLQAY 1057
 Db 594 ELASSLSGRN---ALQDEKRRLEA-----RIALQEELE---EEQGNMEAMSD 635
 QY 1058 EYEVHYVLOEELIDSSPLSDNQRMDKLEKTNSSLRKQNLDLLEQLVANGRIQS----- 1110
 Db 636 RVKATQQAQAEQLSNELATERSTAKNESARQQLERQNKELRSKLHEMEGAVKSKFKSTIA 695
 QY 1111 -LEATI-----EKLSSSEKSKQAMLTLELERSALLQTVVEELRRSA 1151
 Db 696 ALEAKTAQLEEQVEQAREKQAATKSLKQDKKLLKEILLQVEDERKMAEQYKQAEKGNA 755
 QY 1152 EPSDRPECTOPE 1164
 Db 756 RVKQLKRLQEEAE 768

RESULT 13
 US-08-687-080-51
 ; Sequence 51, Application US/08687080
 ; Patent No. 5965427

;; GENERAL INFORMATION:
;; APPLICANT: Gregory Dolganov
;; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
;; NUMBER OF SEQUENCES: 175
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dehlinger & Associates
;; STREET: 350 Cambridge Avenue, Suite 250
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94306
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/687,080
;; FILING DATE: 17-JUL-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/592,126
;; FILING DATE: 26-JAN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sholtz, Charles K.
;; REGISTRATION NUMBER: 38,615
;; REFERENCE/DOCKET NUMBER: 4600-0111.30
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 324-0960
;; TELEFAX: (415) 324-0960
;; INFORMATION FOR SEQ ID NO: 51:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1312 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHEICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.
;; INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 2.78; Score 160.5; DB 2; Length 1312;
Best Local Similarity 18.88; Pred. No. 0.00037;
Matches 175; Conservative 165; Mismatches 366; Indels 223; Gaps 36;

Qy 286 EENRTMLFTIGSEVYLIS-----PDTKKIALEKFKETSCSGIRHVDHGF 334
Db 135 EIDREMISSLGVSKAVLNVIHCHOEDSNWPLSEGKALKQKDFEISATRIKALETLRQ 194
Qy 335 ICRESSGGGPHF---VCYVQCTNEALVDEIMMTLKQAFVAAVQQTAKAPALCEGCP 391
Db 195 V-ROTOGQKVKEYQMLKYLQYKEKACEIRDQITSKEAQLTSSKEIVKSYENEL---DP 250
Qy 392 LOS-----LHKCEIEGNNSSKTKLEQLKHLTTITNOEQATIFEVOKLAPRN 440
Db 251 LKNRLKEIEHNLKIMKLDNIEKALDSRKKOME-----KDNSELEKMEKVFGT 300
Qy 441 EORENELIISFLRCLYEKQK---HHTIGEMKQTSQMAENICSELPSPATRFRLDWLK 497
Db 301 DEQLNDLYHNHQRVREKERKLVDPCHRELEKLNKESRLNNOE-KSELLVEOGRLOLQADR 359
Qy 498 N-----KAKRSUTESLESTLRGNKARG-LOEHSIS-----VDLSDSLSTLSNTSK 543
Db 360 HQEHIRARDSLIQSATOLELDGFERGPFSSRQTKNFHKLVRQEQEAKTANOLMN--- 416
Qy 544 EPSVCEKEAL---PISSEFFLLG-----SSEDLSSDSSESHLPPEAPLSPQOAFRRRA 594
Db 417 --DFAKETLQKQIDEIRDKTGTGRIELKSEILSKKQNELKNVYELQQLGSSDRI 474

Qy 595 NTLSHFPIEQEPPQARGSPGVSQKLMRYHVSYSTETPHERKDFESKANHLGDSGGTPV 654
Db 475 LELDOELIKAEKRELSKAEKNSNVELTKM---EVISLQ--NEKADLDRTLRLKDQE---M 525
Qy 655 KTRRHSWROQIFLRVATPQKACDS-----SSRYED--YSSELGELPPRSPLEPCVEDGPF 706
Db 526 EQLNHHHTTTTQOMELTKDKADKDEQIRKIKSRHSDELTSLLGYFPNKKQLEDWL----- 580
Qy 707 GPPPEEKRTSRRELWQKAILQOILLRMKEKNQKLOASENDL--LNKRLKLDYEBIT 764
Db 581 ----HSSKSEINQTRD-----RLAKLNKELASSEQNKNNHNNELKRKEQL- 622
Qy 765 PCLKEVTVMEKMLSTPGRSKIKFDMEMKHSVAVGGVPRHHRGETWFLAEOFHLKHQFP 824
Db 623 -----SYEDKLFVCGSQDFESDLDRL-----KEETEK-----S 652
Qy 825 SKOOP----KDPYKELIKQLTSQOHAILEDLGRFTTPTHPYFSAQLGAGAGSLYNILKAY 880
Db 653 SKORAMLAGATAVYSQFITQLTDENQSCCPVCQRFQI---EAEI----- 694
Qy 881 SLLDQEVGYCOGLSFVAGILLHNSBEEAFKMLKFLMDFMGLRKQYRDMT----- 931
Db 695 ----QEV-----ISDLSQSKRL-----APDKLKSTESLKKKRRDEMGLVPMRQSI 739
Qy 932 --ILQIQMYQLSRLLHDYHRD---LYNHLREHEIGPSLYAAPWFLTMFASOPLGCVARV 986
Db 740 IDLKEKEIPELKNLQNVNRDIQRLKNDIEQE-----TLTGTIMPEESAKV 787
Qy 987 -----FDMIFLAGTEVIFKVALSLGSHKPLILQHENLETIYDFIKSTLPNLGLV 1036
Db 788 CLTDVTIMERFQEMELKQVERKIAQAAKLGIDLDRTVQOVNQEK-----QEKQHKLDTV 842
Qy 1037 QMKTINQVFEMDIKQLOAYEYHVLOEBELIDSSPLSNQMDKLEKTNSSLRKQNL 1096
Db 843 SSKIELNRKLIQDQEQIQHLKSTNELKSEKQIS--TNLORQLEEQTVELSTEVQS 900
Qy 1097 LLEOLOVANGRIQSLEATIEKLLSSEKLI 1125
Db 901 LYREIKDAKEQVSPLETTLEKFOQEKEEL 929

RESULT 14
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0960
; TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 148:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1312 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54
 US-08-592-126-148

Query Match 2.6%; Score 159.5; DB 2; Length 1312;
 Best Local Similarity 18.8%; Pred. No. 0.00045;
 Matches 175; Conservative 165; Mismatches 366; Indels 223; Gaps 36;

QY 286 EENRTMLFTIGQSEVYLIS-----PDTKIALKKNFKKIFSCQSGIRHVDHFGF 334
 DB 135 EIDREMISSLGVSKAVLNNVIFCHQEDSNWPLSEGLKQKQKDFEIFSATRYIKALETLRQ 194
 QY 335 ICRESSGGGFGF---VCYVFOCTNEALVDEINMTLKQAFVAAVQOTAKAPACQCEGCP 391
 DB 195 V-RTOGQVKVQOMELYLKQKACEIRDQITSKEAQLTSSKEIVKSYENEL---DP 250
 QY 392 LQS-----LHKLCRIEIGMSSSKTKLEQLKHLTLTNOEQATIFEVEYQKLRPRN 440
 DB 251 LKNLKEIEHNLSKIMKLDNEIKALDSRKKQMB-----KDNSELEKMEKVFQGT 300
 QY 441 EQRENELISFLCLVEEKOKE---HIHIGEMKQTSOMAENIGSELPPSATRFRDLMLK 497
 DB 301 DEQLNDLYNHQRTVREKRLVDCHRELEKLNKESRLNQE-KSELLVEQGRLOQADR 359
 QY 498 N-----KAKRSLTSLSLSISLGNKARG-LQEHIS-----VLDSSLSLTLSNTSK 543
 DB 360 HQHIRARDSLQSLATQLELDFGFGPFSERQIKNFKHLVRERQGEAKTANQLMN--- 416
 QY 544 EPSVCEKEAL---PISESFKLIG-----SSDLSSESHPLEPEAPLSPQAFRRRA 594
 DB 417 -DFAKETLQKQIDRDKTKGLRIETELKSEILSKQNELKNVYELQLEGSSDRI 474
 QY 595 NTLSHFPICQEPQPARSGPVGSRKLMRYHSVSTETHERKDPESKANHLGDSOGTPV 654
 DB 475 LEDQLELIKAEBSLKAENSNVETLKM-----EVLISQ--NEKADLORTLRKLDQE---M 525
 QY 655 KTRHRSWROQIFURVATPOKACDS-----SSRYED--YSELGELPPSPLEPVCEGDF 706
 DB 526 EQLNHHITTTQEMETKQADKDEQIRKIKSRHSDELTLGYPFNKKQLEDWL----- 580
 QY 707 GPPPEKKRTSRELRLNOKAILQQILLRMEKENOKLASENDL--LNKRLKLDYEIT 764
 DB 581 ---HSKSKEINQTRD-----RLAKLNKELASSEQNKNHNNELKRREQL- 622
 QY 765 PCLKEVTVWEKMLSTPGRSKIKFDMKMHSAVGQGVPRHHRGEIMKFLAEQPHLKQHP 824
 DB 623 -----SSYEDKLFVCGSODFSDDLRL-----KEIEK-----S 652
 QY 825 SKQOP-----KDVVPYKELLKQITQQHAILIDLGRTPPTHPYFSAQLGAGOLSLYLKAY 880
 DB 653 SKQRAMLAGATAVYSQFITQIDENQSCPCVQCVFTQ-----EAL- 694
 QY 881 SLLDQEVGYCGSLFVAGILLMLHMBEEAFKMLKFLMDFMGLKRYRQPMI----- 931
 DB 695 ---QEV-----ISDLQSKRL-----APDKLKSTESLKKKREKRDMLGLVPMRQSI 739
 QY 932 --ILQIQMYOLSRLLHDYHRD--LYNHLREHEIGFSLYAAPWFLTMFAQSOPFLGVFARV 986
 DB 740 IDLKEKEIPELRNKLQNVNDIQLKNDIEQE-----TLTGTMPEESAKV 787
 QY 987 -----FDMIFLQGVTEVIFKVALSLGSHKPLILOHENLETIVDFIKSTPLNLGLV 1036
 DB 788 CLTDVTIMERFQMLKDVVERKIAQQAALQGLIDLDRTVQVQVNEK-----QEKQKLDTV 842
 QY 1037 OMEKTIINQVFEMDIQALQAYEYEHVQLQEBLIDSSPLSNQRMKLEKTNSSLRKQNL 1096

DB 843 SSKIELNRRLIQDQOQIQIHLKSTTNELSEKLOIS--TNLQRRQQLSEOTVELSTEVS 900
 QY 1097 LLEQLOVANGRIQSLSEATIEKLSSSKL 1125
 DB 901 LYREINKDAKEQVSPLETTLEKTFQKEEL 929

RESULT 15
 US-09-060-410-2
 ; Sequence 2, Application US/09060410
 ; Patent No. 6165461
 ; GENERAL INFORMATION:
 ; APPLICANT: Cobb, Melanie
 ; APPLICANT: Hutchinson, Michele
 ; APPLICANT: Chen, Zhu
 ; APPLICANT: Berman, Kevin
 ; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
 ; TITLE OF INVENTION: THEREFOR
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/060,410
 ; FILING DATE: 14-APR-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 860098.421
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1001 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-060-410-2

Query Match 2.6%; Score 156.5; DB 4; Length 1001;
 Best Local Similarity 18.8%; Pred. No. 0.00052;
 Matches 153; Conservative 111; Mismatches 282; Indels 267; Gaps 32;

QY 410 TKLELQKHHTLTNQDQATIFEVEYQKLRPRNEQRENELISFLRCLY-----EE 458
 DB 271 TSELKLMFVLRPERPETVLIDLQRTKDAVRELDNLQYRKMKKLLFQEAHNGPAYEAQE 330
 QY 459 KQKEHTHIGEMKQTSOMAENIGSELPPSATRFRDLMLNKAKRSLTESLTSRGNKA 518
 DB 331 EEEQDHGGGRTGT-----VNSVGS-----NQSIPTSM----- 357
 QY 519 RGLQEHISVDLSLSTLSNTSKPSVCEKALPISESKLGSSEDLSSDSSEHL- 577
 DB 358 -----SISASSQSSVNSLPOASD-----DKSELDMM-----GDHTVMSNSSVIHLK 400
 QY 578 PEB-----PAPLSPOQAFFRRAN-----TLSHFPTECQPPQPARC 613
 DB 401 PEBENTYQEGDPRTRASAPQSPQVSRHKSHYRNREHFATIRTAASLVTRQMOHEQDSEL 460
 QY 614 SPQVSQRLMR-----YHSVSTETPHER-----KDFESKANHLGDSGGTPVKTTRH 659

```

Db 461 REQSGYKRMRRHQKQKQMTLENKLAEMDEHRLDKDLETOQNNFAAEMKLIKXQA 520
QY 660 SNRQIFLRVATPQACDSSSYDYSELGELPPSPCEDGPGPPPEEKRTSRE 719
Db 521 SMEKAHV-MANEERKFOQHIQAQOKKELNSFLESQKREYKLRKEQLKEELNENQSTPKK 579
QY 720 LRELKQATLQOILLRMEKEN-QKLOASENOLLNKRKLKDYEEITPCLKEVTTVWEKML 778
Db 580 EKQEN-----LSQKENIQHFAEEANLLRROR-QYLEL-----ECRRFRKRWL 623
QY 779 STPGRSKIKFDMKMHSAVGCQVPRHRGEIWKFLAEQFHLKHOPPSKQPKDVPYKELL 838
Db 624 -----LGRHN-----LEQDLVREELNKRQTKQKDLSEHAMLL 653
QY 839 KQITSQO-----HAILIDGRTPFPHPYFSAQLGAGOLSLYNILKAYSLLDOEVCYCOGL 893
Db 654 ROHESMQELEFRHLNTIQMRCELIRLQHQTEL-TNOLE-YNKRRELERKRV----- 705
QY 894 SFVAGILLIHMSEEAFAKMLKFLMDFMGLRKOYRPDMITLQIQMYQLSRLLDHYHRLYN 953
Db 706 -----NEVROQPKSLK--SKELQIKKQFQ-DTCKIQTRQYKALR-----N 742
QY 954 HLEEHEIGPSLYAAPWFLTMFASQPLGFVARVDMIFLQGTVEIFKVALSLGSHKPLI 1013
Db 743 HLE-----TPKSEHKAVL 757
QY 1014 --LQHENLETIVDFIKSLPNIGLVOMEKTIQVFMEDIKQLOQAYEVEYHVLOEELIDS 1071
Db 758 KRLKEEQTRKLAIAE-----QYDHSINEMLSQALRLDEAQAEQVLMQOL--- 805
QY 1072 SPLSDNORMDKLEKTNSSLRQ-----NLDLLEQLQVANGRIQSLQSLATIEKLSSSEK 1124
Db 806 -----QQELELLNAYQSKIKQAEQAQHDRELRELEQ-----RVSRLRALLEQ-----K 848
QY 1125 LKQAMLTLELERSALLQVTEELRRRSAPSDRE 1157
Db 849 IEEEMALQONERTIRSLERQAREIEAFDSE 881

```

Search completed: August 28, 2002, 16:23:08
 Job time: 41 sec